

THIS WEEK



EDITORIALS

WORLD VIEW The ungainly snowball fight that is university metrics p.255

SAND Smashing grain collisions drive severe desert storms p.256

BEHAVIOR Demonstrator skill helps birds get the worm p.257

Beyond headscarf symbolism

Turkey's scientists show they no longer want to expend their energy on political confrontation, but political volatility is threatening their efforts to work in peace.

This spring in Istanbul, a hundred or so faculty members and students held a lunch-time demonstration in one of Koç University's elegant courtyards. They were protesting against the administration's failure to protect the jobs of the university's service workers. It was peaceful. No one tried to stop them.

This is one side of Turkey. The other side is more visible — the violent response of police to the Taksim Square protests against unchecked city development a few weeks later, for example. And the long jail sentences dished out earlier this month to some former university rectors, dogmatic opponents of their country's ban on headscarves in public institutions (see *Nature* 500, 129–130; 2013). The headscarf ban, and the reaction to it, is a telling and useful guide to how Turkish universities and Turkish science could yet flourish amid such political volatility.

Koç University is private, so the ban does not apply. Free to choose, only a single head was covered among the protestors. With reasonable budgets, reasonable student numbers and a reasonable absence of ideology, Koç University has little to worry about. Its three-year-old medical faculty has already attracted a European Research Council grant holder.

A few kilometres farther south, the cash-starved, public Boğaziçi University is required to apply the headscarf ban, but the occasional defiant student who chooses to cover her head tends to go unnoticed. As elsewhere in Turkey, most academics and scientists just don't want to expend energy on fights about headscarves any more.

That is a U-turn in mentality, and an encouraging one. The headscarf ban has been the symbol of a decades-long struggle between secularists who see covering the hair as a slippery slope towards a breakdown of Turkey's constitutional separation of state and religion, and those who would like to see a greater presence of Islam in daily life.

Scientists tend increasingly to see this stand-off for what it is — less about religion than about power, and they are worried more about the ascension of ignorance than of fundamentalism.

Turkey's economic growth and expanding cities have created a powerful new middle class with strong religious roots. Ten years ago they helped to elect Prime Minister Recep Tayyip Erdogan with his 'mildly Islamic' government, and Erdogan supporters now hold key positions at institutions such as the Turkish Academy and TÜBİTAK, Turkey's research funding agency. Critics say that these individuals are not always qualified for the posts, and that some may even have abused their positions to take revenge on those who stood in their way in the past.

TÜBİTAK has made a series of poor decisions in recent years. Most notorious was in 2009, when it demoted Çigdem Atakuman, the editor of its magazine *Bilem ve Teknik*, after she objected to its censorship of an article celebrating Charles Darwin. She has since challenged TÜBİTAK's investigations in three separate court cases, which she won. More recently, the organization declined to fund a workshop on evolutionary biology, saying that evolution is controversial. Anti-evolutionism is a religious stance, and with this argument TÜBİTAK shows a failure to understand science and its processes.

Against this alarming backdrop, Turkey's small, mostly foreign-trained research community is trying to do science as it would be done in the countries where its members trained. They frequently succeed in small pockets, as at Koç and Boğaziçi. But even at Koç, scientists complain about a lack of critical mass that threatens their long-term future. The current less-confrontational mood will help them to focus their energy. But they won't be secure until the government acknowledges that science, just like a headscarf, is not a political toy. ■

In addition

Conflicts of interest and gaps in data contaminate US oversight of food additives.

A chef who crafts a delicacy for sale in the United States can choose from more than 10,000 food additives to garnish the dish. Of these chemicals, 43% are labelled 'generally recognized as safe' (GRAS) and need not be approved by the US Food and Drug Administration (FDA).

The system has weaknesses. A manufacturer is responsible for assessing whether an additive it has developed is GRAS. Once that is done, the manufacturer is asked — but not required — to notify the FDA. There are no data to evaluate compliance systematically, but the FDA found during a 2010 crackdown on caffeinated alcoholic drinks that four out

of four manufacturers queried had not done the required checks.

Even when manufacturers do submit GRAS determinations, there are concerns about the quality of the assessment. An ongoing project at the Pew Charitable Trusts in Washington DC reveals discomfiting gaps in the data. A search of three toxicological databases, including that of the FDA, showed that fewer than 38% of GRAS claims were backed up by FDA-recommended toxicology studies in animals (T. G. Neltner *et al.* *Reprod. Toxicol.* <http://dx.doi.org/10.1016/j.reprotox.2013.07.023>; 2013). The same team has published an analysis of 451 GRAS notifications submitted to the FDA. To avoid conflicts of interest, assessments should be done by an independent expert panel, but none had been; in fact, 22% had been performed by an employee of the manufacturer (T. G. Neltner *et al.* *JAMA Intern. Med.* <http://doi.org/nd5>; 2013).

The FDA told *Nature* this week that it plans to issue guidance on how to fulfil GRAS requirements. One suggestion, proposed by Pew, is to take conflict-of-interest policies used to select FDA advisers, and apply them to GRAS determinations. The FDA should seize the opportunity to protect public health and boost confidence. ■



Halt the avalanche of performance metrics

The increasing dominance of quantitative research assessment threatens the subjective values that really matter in academia, says Colin Macilwain.

The leaders of major universities around the world used to maintain a healthy scepticism towards league tables and the metrics that underpin them. But now, officials at institutions that do well in such assessments — partly on merit, and partly because they use the English language or have other historical advantages — are becoming beguiled with quantitative measures to rate the performance of academic staff. People who care about genuine quality in research and teaching need to resist that shift.

Universities evolved as self-governing bodies of academics. Originally, the president or vice-chancellor had a purely housekeeping role, once described in US parlance as assuring parking for the staff, sex for the students and sport for the alumni.

But lately — not least in Britain, where schemes such as the Research Assessment Exercise have come to dominate academic life — power has moved from the departments to the vice-chancellor. And university leaders, flanked by research managers and associated flunkies, want to use metrics to shift that balance still further.

Eight leading British universities are now energetically engaged in the joint development of a formidable computer tool that allows them to compare the performances of their researchers and departments against rivals, according to grant income, number of patents applied for, or pretty much any other criteria they choose. The tool is called Snowball (www.snowballmetrics.com) and the institutions signed up to it include the universities of Oxford and Cambridge, Imperial College London and University College London.

Like any metrics system, Snowball can, in theory, be used for good or ill. I suspect that in practice, however, it will end up being used mainly to exercise yet more control over academic staff, with every aspect of their professional lives tracked on the system.

Although Snowball was developed by people of genuine integrity who want to establish a fuller understanding of research performance, it shares a fundamental defect with other quantitative research-assessment tools: it is largely built on sand. It cannot directly measure the quality of research, never mind teaching, so instead it uses weak surrogates, such as the citation indices of individuals.

Citation indices — which rank research in terms of the average number of citations for articles — were robustly challenged earlier this year, when organizations led by the American Society for Cell Biology signed the San Francisco Declaration On Research Assessment (DORA), pledging to take a stand against the ever-expanding reach of journal-based metrics.

One of DORA's best ideas is to ask that citation databanks be available openly, for all researchers to use. I wish them luck with that. University

managers know that information is power — and they want not just the data, but to dictate how they are manipulated.

A major problem with metrics is the well-charted tendency for people to distort their own behaviour to optimize whatever is being measured (such as publications in highly cited journals) at the expense of what is not (such as careful teaching). Snowball is supposed to get around that by measuring many different things at once. Yet it cannot quantify the attributes that society values most in a university researcher — originality of thinking and the ability to nurture students. Which is not the same as scoring highly in increasingly ubiquitous student questionnaires.

Senior scientists have known for a long time that bogus measures of 'scientific quality' can threaten the peer-review system that has been painstakingly built up, in most leading scientific nations, to distribute funds on the basis of merit.

In the United States in 1993, for example, Congress passed the Government Performance and Results Act, which compelled federal agencies to start measuring their results. However, the US scientific establishment was strong and self-assured at that time, and successfully derailed the prospect that agencies such as the National Science Foundation (NSF) would start inventing numbers to 'measure' the work of its grant recipients. Instead, the NSF sticks to measuring things such as time to grant.

Nations with weaker scientific communities are less well-placed to fend off the march of metrics. The hazards are perhaps most immediate in places such as Italy, where peer review for grants has never fully taken hold, and China, where it has rarely even been tried. There is a worrying tendency in developing countries, especially, for research agencies to skip the nuanced business of orchestrating proper peer review, and to move straight to the crude allocation of funds on the basis of measured performance. This bypasses quality and, bluntly, invites corruption.

But I see trouble ahead at the leading universities in the United Kingdom and the United States, too. Their reputations were built by autonomous academics, working patiently with students. If the name of the game becomes strong performance measured in numbers — as the vice-chancellors seem to want — it will kill the goose that laid the golden egg.

Defenders of Snowball say they are baffled that scientists, given what they do for a living, remain sceptical of research-performance metrics. But science seeks to identify and measure good surrogates, to test falsifiable hypotheses. Seen in that light, quantifiable research assessment does not measure up. Nevertheless, the snowball has started rolling down the mountain — and it is hard to see how its momentum will be arrested. ■

BOGUS MEASURES OF **'SCIENTIFIC QUALITY'** CAN THREATEN THE **PEER-REVIEW SYSTEM.**

► NATURE.COM

Discuss this article
online at:
go.nature.com/fpfpwg

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfnworldview@gmail.com

RESEARCH HIGHLIGHTS

Selections from the scientific literature

ENVIRONMENTAL SCIENCES

Sinking ground poisons wells

A previously unrecognized source of arsenic threatens Vietnam's clean water.

Arsenic contamination generally affects wells at shallow depths, so wells deeper than 150 metres are often deemed safer. However, Steven Gorelick and colleagues at Stanford University in California suggest how arsenic contamination can occur deep underground. They analysed an area in the Mekong Delta of Vietnam, where many deep wells are contaminated. Using satellite measurements and simulations, they found that the land around deep wells has sunk by up to 27 centimetres since 1988 as groundwater has been pumped out. Clay layers adjacent to the pumped areas have compacted, which could be forcing water containing dissolved arsenic into deep aquifers. Wells that initially test as arsenic-free may not remain so, the authors say.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1300503110> (2013)

BIOMECHANICS

Spider silk stalls a somersault

A silk string helps leaping spiders to land fast on their feet.

Jumping spiders, or salticids, can travel more than 20 times their body length in a single bound, usually with a dragline of silk, thought to function as a safety line, trailing behind them. Kai-Jung Chi at National Chung Hsing University in



PAUL KENNEDY/LONELY PLANET IMAGES/GETTY

GEOPHYSICS

Sand collisions kick up storms

Sandstorms whipped up by desert winds owe their strength to mid-air collisions between sand grains.

Marcus Carneiro and Hans Herrmann at the Swiss Federal Institute of Technology in Zurich and their colleagues used computer simulations to show that without collisions, sand grains whizz along close to the ground or make small hops. But when they encounter each other in the

air, a few particles are bounced upwards off their neighbours and are accelerated by the stronger winds that blow higher up. These high-flying 'saltans' ultimately crash into the ground, kicking out sprays of sand that further fuel the storm.

Such collisions can double the amount of material in the air, boosting the quantities that storms carry as they sculpt desert landscapes.

Phys. Rev. Lett. 111, 058001 (2013)

Taiwan and her colleagues showed that the dragline also provides stability, preventing the predatory spiders from pitching too far back through the air and so leaving them poised for action on landing.

Time-lapse images of the salticid *Hasarius adansoni* showed that spiders with silk (**pictured; left panels**) maintained orientation during jumps: they landed feet-first and were ready to pounce within about 10 milliseconds. Salticids that did not produce silk (right panels) landed on their abdomens, slipped or tumbled, sometimes requiring

more than 50 milliseconds to regain their footing.

J. R. Soc. Interface 10, 20130572 (2013)

CLIMATE SCIENCE

Climate tracking by smartphone

Scientists can gather climate data from smartphones in their owners' pockets.

Aart Overeem of Wageningen University, the Netherlands, and his colleagues exploited an existing Android smartphone app to collect

and select around 1.3 million temperature readings from eight cities around the world, averaging more than 800 readings per city each day. The app gathered data from phone-battery temperature sensors and uploaded the readings when a data connection was available. A simple heat-transfer model accounted for the effects of body warmth and clothing, and allowed the researchers to estimate daily temperatures in urban environments. The approach could be used to collect data at a finer scale and lower cost than is currently possible with weather stations, the authors suggest.

Geophys. Res. Lett. <http://dx.doi.org/10.1002/grl.50786> (2013)

256 | NATURE | VOL 500 | 15 AUGUST 2013

© 2013 Macmillan Publishers Limited. All rights reserved

YUNG-KANG CHEN; KAI-JUNG CHI/JRSI

PHYSIOLOGY

Why sunburn stings

A membrane protein in skin cells transmits signals that trigger pain and tissue damage that put the 'burn' into sunburn.

Wolfgang Liedtke of Duke University in Durham, North Carolina, and his colleagues switched off production of a cellular ion channel called TRPV4 in the epidermis of genetically engineered mice. They exposed the mice to ultraviolet (UV) radiation, which causes sunburn.

Compared with UV-exposed mice that had normal TRPV4 production, the engineered mice had less damage on their hind-paw pads (which resemble human skin). Their paws were also less sensitive to heat or to being poked with fine wires — signs that the engineered mice experienced less pain. Furthermore, three human patients with sunburn had higher expression of TRPV4 in their epidermis than three unburned controls.

Proc. Natl Acad. Sci. USA
<http://dx.doi.org/10.1073/pnas.1312933110> (2013)

TROPICAL DISEASE

Malaria vaccine protective at last

A vaccine made of a purified, weakened form of a malaria parasite gave 100% protection in a small clinical trial.

Of six people who were given five intravenous doses of the live-attenuated vaccine and then bitten by infectious mosquitoes, none developed malaria. Five out of six unvaccinated people did get the disease, as did three out of nine patients who received lower doses of vaccine. These levels of protection, reported by Robert Seder at the National Institutes of Health in Bethesda, Maryland, and his colleagues, surpass any previously achieved for a malaria vaccine. The vaccine will now be tested in larger field trials in Africa,

which will determine how long protection lasts and whether it works against other strains of the malaria parasite.

Science <http://dx.doi.org/10.1126/science.1241800> (2013)

For a longer story on this research, see go.nature.com/mae5tu

PALAEONTOLOGY

Toothy patch in fossil fish

Jawless fish that swam the seas some 400 million years ago may have relied on the same mechanism to repair bone damage as modern mammals do to mend their teeth.

A team led by Zerina Johanson at the Natural History Museum in London analysed fossils of the large fish *Psammolepis* from rock formations in Estonia and Latvia. *Psammolepis* are covered by bony shields ornamented with a thin layer of dentine, the same tissue that makes up teeth. One specimen had a damaged armour plate, possibly from a predator attack. The researchers were surprised to find that the bony plate was plugged not with bone, but with dentine.

Specialized stem cells speed the regrowth of dentine in extant mammals such as humans and mice, and the researchers suggest that a similar mechanism existed in this toothless fish to repair its bony armour.

Biol. Lett. <http://dx.doi.org/10.1098/rsbl.2013.0144> (2013)

CANCER BIOLOGY

Genome mix-ups monitored

Cancer cells often contain translocation mutations, in which parts of the genome are swapped between chromosomes. For the first time, scientists have been able to watch these accidents as they happen in living cells.

Researchers led by Tom Misteli at the National Cancer Institute in Bethesda, Maryland, and his colleagues, coupled

COMMUNITY CHOICE

The most viewed papers in science

ANIMAL BEHAVIOUR

Young female birds get the worm

HIGHLY READ

on www.elsevier.com in the past 90 days

Birds' ability to learn from watching their peers varies with age, sex and social rank. In a study led by Lucy Aplin at the University of Oxford, UK, wild-caught blue tits (*Cyanistes caeruleus*; famous for learning to pierce foil caps on milk bottles delivered to British homes to get the cream) were placed in an aviary with a covered tray containing waxworms, the blue tit's favourite snack. Eight experimental groups totalling 56 birds could observe 'demonstrator' birds that had been trained to get worms by either flipping up or piercing the covers. About half of these birds learned to get the worms, but when 32 other birds were given the tray in the absence of demonstrators, none of them learned the skill. Juvenile females were almost twice as likely to learn the task as other blue tits, and dominant males less likely.

Anim. Behav. 85, 1225–1232 (2013)



MARCELIB/Flickr/GETTY

ultra-high-throughput techniques for time-lapse imaging with a system for 'snapping' DNA strands in chromosomes, and watched what happened to DNA fragments in cultured cells. On rare occasions, the breaks occurred in separate chromosomes and the broken ends attached to the wrong partners. These translocations formed within hours of a break and were guided by the cell's repair machinery. The work will allow further study of how and when chromosomal mix-ups occur.

Science 341, 660–664 (2013)

REPRODUCTION

Sneaky breeders make sons

Altruistic worker bees usually abstain from reproduction — unless they have a shot at making a royal baby.

Michael Holmes at the University of Sydney in Australia and his colleagues

monitored the fertility and breeding habits of seven colonies of Western honeybees (*Apis mellifera*). It was thought that workers (who are all female) refrain from breeding and instead raise the queen's brood, which is better for the colony's overall success. But the researchers found that overall, 4.2% of males (drones) are produced by workers, a level 40 times greater than that typically reported. Furthermore, just before virgin queens set out to establish their own hives, a worker baby boom occurs and the number of worker-produced drones peaks at more than 6%.

The authors speculate that workers may risk breaking rank for the chance of producing a new queen's future mate.

Mol. Ecol. 22, 4298–4306 (2013)

NATURE.COM

For the latest research published by *Nature* visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

RESEARCH

Coronavirus clues

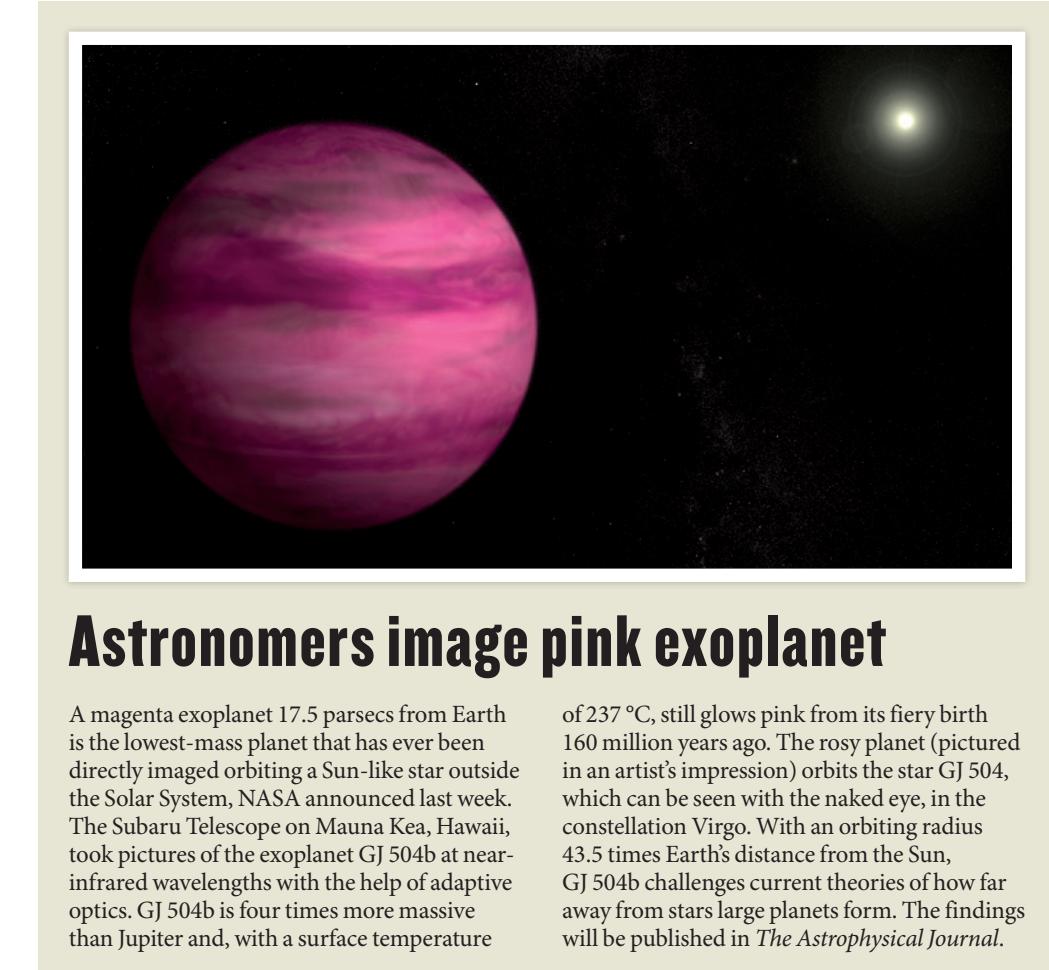
Scientists have an early lead in the search for animal sources of the Middle East respiratory syndrome coronavirus (MERS-CoV), which has killed 46 people since last September. The virus is thought to have originated in bats, but could have spread to humans through one or more intermediate hosts. Researchers reported on 9 August that of 50 Omani camels sampled, all showed evidence of previous infection with MERS-CoV or a closely related virus (C. B. E. M. Reusken *et al.* *Lancet Infect. Dis.* <http://doi.org/nfb>; 2013). Further work is needed to verify whether camels could be a source of human infections.

Stormy Atlantic

The current Atlantic hurricane season, which began on 1 June, could be unusually active, said the US National Oceanic and Atmospheric Administration in an updated forecast on 8 August. There have been four named storms so far this year. The agency calculates a 70% chance that a total of 13–19 named storms, including 3–5 major hurricanes, will develop before the Atlantic season ends on 30 November. These projections would exceed the 30-year seasonal average of 12 named storms, including 3 major hurricanes.

GM rice row

Protesters uprooted a field of genetically modified (GM) 'golden rice' at a Philippines Department of Agriculture compound in Camarines Sur on 8 August, citing concerns over future marketing and potential health effects. The crop is engineered to contain β-carotene and helps to



S. WEISSINGER/NASA GODDARD SPACE FLIGHT CENTER

Astronomers image pink exoplanet

A magenta exoplanet 17.5 parsecs from Earth is the lowest-mass planet that has ever been directly imaged orbiting a Sun-like star outside the Solar System, NASA announced last week. The Subaru Telescope on Mauna Kea, Hawaii, took pictures of the exoplanet GJ 504b at near-infrared wavelengths with the help of adaptive optics. GJ 504b is four times more massive than Jupiter and, with a surface temperature

of 237 °C, still glows pink from its fiery birth 160 million years ago. The rosy planet (pictured in an artist's impression) orbits the star GJ 504, which can be seen with the naked eye, in the constellation Virgo. With an orbiting radius 43.5 times Earth's distance from the Sun, GJ 504b challenges current theories of how far away from stars large planets form. The findings will be published in *The Astrophysical Journal*.

mitigate vitamin A deficiency — which causes malnutrition and affects 1.7 million children in the Philippines. The non-profit International Rice Research Institute, which was helping to conduct safety trials on the rice, vowed to continue its experiments.

Landsat woes

The future of Landsat, the world's longest-running series of Earth-observing satellites, is in jeopardy, according to a report released on 8 August by the US National Research Council. Current methods of management and funding cannot sustain the land-imaging programme over the long term, the report warns. NASA launched the

US\$855-million Landsat 8 probe in February (see *Nature* **494**, 13–14; 2013), but the space agency has not yet outlined clear mission objectives or secured sufficient funding for a successor. Landsat satellites have monitored Earth's surface continuously since 1972.

H7N9 virus persists

China reported on 11 August its first new case of the H7N9 avian influenza virus in three weeks: a 51-year-old woman from Huizhou in Guangdong province, who fell ill on 27 July. Since the start of June, only three cases have been confirmed in China; 132 cases were confirmed from March to May. There have been

44 deaths, but no evidence of sustained human-to-human transmission, says the World Health Organization.

BUSINESS

Brain-implant trial

A deep-brain stimulation (DBS) device that not only delivers electrical pulses, but also records brain activity simultaneously, has been implanted in a person for the first time. Until now, there has been a lack of data on how the brain responds to DBS, which is used to treat motor disorders such as Parkinson's disease, and is being tested for some psychiatric conditions. On 7 August, the device maker — Medtronic in Minneapolis,

PAULE ALERS/NASA Minnesota — announced the start of clinical trials for the system, with the first implantation in a person with Parkinson's in Germany.

Genetics lawsuit

Ambry Genetics in Aliso Viejo, California, has countersued its competitor Myriad Genetics. Myriad, a medical diagnostics company in Salt Lake City, Utah, sued Ambry in July for infringing patents that Myriad holds on tests for cancer-associated mutations in the *BRCA1* and *BRCA2* genes. Last week, Ambry replied that Myriad's patent claims were invalid in light of a June ruling by the US Supreme Court that human genes cannot be patented (see *Nature* 498, 281–282; 2013), and countersued Myriad for suppressing competition to maintain a monopoly on the tests.

PEOPLE

NASA deputy out

NASA deputy administrator Lori Garver (pictured) is resigning after four years, the agency announced on 6 August. As second in command to NASA chief Charles Bolden, Garver had a major role in developing initiatives such as commercial space transportation and a mission to capture an



asteroid. She has also advised US President Barack Obama on space policy during his campaign and presidency. Garver will become general manager at the Air Line Pilots Association, based in Washington DC.

Misconduct finding

A dermatology researcher at Case Western Reserve University in Cleveland, Ohio, has been sanctioned for misconduct by the US Office of Research Integrity (ORI). Pratima Karnik admitted submitting a grant application to the US National Institutes of Health that contained text plagiarized from a grant application she had reviewed, as well as from eight research articles and a patent application. The ORI, in Rockville, Maryland, noted on 6 August that Karnik has agreed to have her research

supervised, and to exclude herself from any peer-reviewing or advising for US public-health services for two years.

POLICY

US biofuel quotas

In a long-awaited decision, on 6 August the US Environmental Protection Agency (EPA) announced the nation's biofuel quotas for 2013. Biofuels — mostly those based on ethanol from maize (corn) — must account for at least 9.74% of the total US fuel supply this year, up from 9.23% in 2012. However, the percentage of cellulosic biofuels — those made from inedible plant fibres such as wood — was reduced from last year's target of 0.006% to 0.004% after a federal court ruled in January that the fledgling industry would not be able to meet the EPA's demands.

Food safety concern

Additives that help to colour, flavour and preserve food are often officially deemed safe by experts with financial ties to the manufacturers, according to data published on 7 August (T. G. Neltner *et al.* *JAMA Intern. Med.* <http://doi.org/nd5>; 2013). The study, led by researchers at the Pew Charitable Trusts in Washington DC, found that

COMING UP

20 AUGUST

A US presidential bioethics commission discusses the ethical implications of neuroscience and its applications at a public meeting in Philadelphia, Pennsylvania (see page 261). go.nature.com/zrthim

22 AUGUST

US federal agencies reach the deadline to file their proposals for increasing public access to federally funded research, set by the Office of Science and Technology Policy in Washington DC (see *Nature* 494, 414–415; 2013).

employees or hired consultants of food companies made 35% of the 451 determinations classifying food additives as safe in the United States between 1997 and 2012. The authors have called on the US Food and Drug Administration to bar the practice. See go.nature.com/jqccur and page 253 for more.

Space fence down

Citing the US federal budget cuts known as sequestration, the US Air Force Space Command announced on 12 August that it will shut down a key component of a network that monitors orbital debris and space objects. The Air Force Space Surveillance System, dubbed the space fence, will close by 1 October. Two of the system's nine radar stations were deactivated in April. Losing the rest will make it harder to track break-ups of space junk, which can endanger satellites, says Brian Weeden, technical adviser at the Secure World Foundation in Washington DC.

NATURE.COM

For daily news updates see:
www.nature.com/news

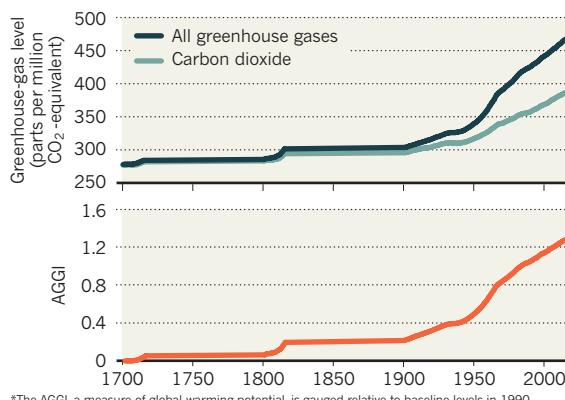
SOURCE: NOAA

TREND WATCH

Increased emissions of carbon dioxide, nitrous oxide and methane drove atmospheric greenhouse gases to their highest recorded levels in 2012, according to the US National Oceanic and Atmospheric Administration (NOAA). CO₂ now accounts for about 80% of the increase in global-warming potential since 1990, the agency says. In May, atmospheric concentrations of the gas surpassed 400 parts per million at NOAA's Mauna Loa Observatory in Hawaii (see *Nature* 497, 13–14; 2013).

RECORD MEASUREMENT OF GREENHOUSE GASES

The US National Oceanic and Atmospheric Administration's Annual Greenhouse Gas Index (AGGI*) rose by 1.5% last year.



NEWS IN FOCUS

RUSSIA Hopes pinned on elite tech campus to fight brain drain **p.262**

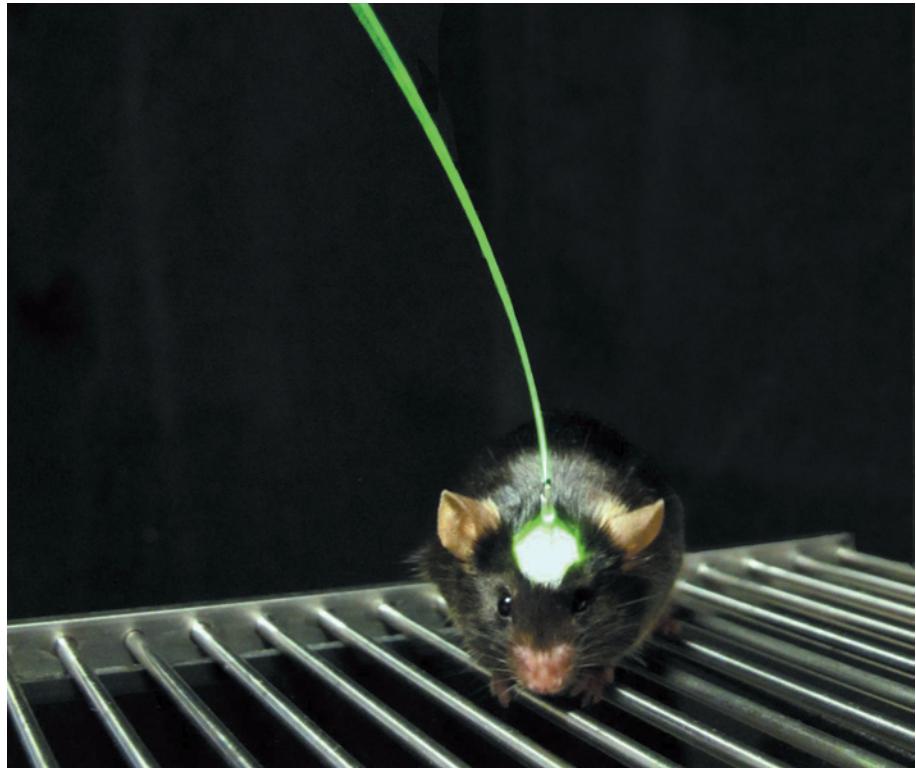
PHYSICS Romania dreams of 'laser valley' to light up economy and science **p.264**

PHARMACEUTICALS India takes a stand over cancer-drug patents **p.266**



PSYCHOLOGY Elizabeth Loftus, maven of malleable memory **p.268**

INBAL GOSHEN & KARL DEISSEROTH



Optical stimulation of light-responsive neurons in engineered mice can be used to create false memories.

NEUROSCIENCE

US brain project puts focus on ethics

Unsettling research advances bring neuroethics to the fore.

BY HELEN SHEN

The false mouse memories made the ethicists uneasy. By stimulating certain neurons in the hippocampus, Susumu Tonegawa and his colleagues caused mice to recall receiving foot shocks in a setting in which none had occurred¹. Tonegawa, a neuroscientist at the Massachusetts Institute of Technology in Cambridge, says that he has no plans to ever implant false memories into humans — the study, published last month, was designed

just to offer insight into memory formation.

But the experiment has nonetheless alarmed some neuroethicists. "That was a bell-ringer, the idea that you can manipulate the brain to control the mind," says James Giordano, chief of neuroethics studies at Georgetown University in Washington DC. He says that the study is one of many raising ethical concerns, and more are sure to come as an ambitious, multi-year US effort to parse the human brain gets under way.

The BRAIN (Brain Research through

Advancing Innovative Neurotechnologies) Initiative will develop technologies to understand how the brain's billions of neurons work together to produce thought, emotion, movement and memory. But, along with the discoveries, it could force scientists and society to grapple with a laundry list of ethical issues: the responsible use of cognitive-enhancement devices, the protection of personal neural data, the prediction of untreatable neurodegenerative diseases and the assessment of criminal responsibility through brain scanning.

On 20 August, US President Barack Obama's commission on bioethics will hold a meeting in Philadelphia, Pennsylvania, to begin to craft a set of ethics standards to guide the BRAIN project. There is already one major mechanism for ethical oversight in US research: institutional review boards, which must approve any studies involving human subjects. But many ethicists say that as neuroscience discoveries creep beyond laboratory walls into the marketplace and the courtroom, more comprehensive oversight is needed. "The long-term consequences of more brain knowledge — whether it's good for an ethnic group or threatens your personal identity — there's sort of no one in charge of that," says Arthur Caplan, director of medical ethics at New York University's Langone Medical Center.

Tonegawa's study adds to the growing evidence that memories are surprisingly pliable. In the past few years, researchers have shown that drugs can erase fearful memories² or disrupt alcoholic cravings in rodents³. Some scientists have even shown that they can introduce rudimentary forms of learning during sleep in humans⁴. Giordano says that dystopian fears of complete human mind control are overblown. But more limited manipulations may not be far off: the US Defense Advanced Research Projects Agency (DARPA), one of three government partners in the BRAIN Initiative, is working towards 'memory prosthetic' devices to help soldiers with brain injuries to regain lost cognitive skills.

Deep brain stimulation (DBS), in which implants deliver simple electrical pulses, is another area that concerns neuroethicists. The devices have been used since the 1990s to treat motor disorders such as Parkinson's disease, and are now being tested in patients with psychiatric conditions such as obsessive-compulsive disorder and major depression. Giordano says that applying DBS technology more widely ►

► requires ethical care. "We're dealing with things affecting thought, emotion, behaviour — what people hold valuable as the essence of the self," he says.

Neuroethicists are noticing challenges beyond the medical system, too, particularly in the courtroom. Judy Illes, a neurology researcher at the University of British Columbia in Vancouver, Canada, and co-founder of the International Neuroethics Society, says that brain imaging could affect the criminal-justice system by changing definitions of personal responsibility. Patterns of brain activity have already been used in some courtrooms to assess the mental fitness of the accused. Some ethicists worry that an advanced ability to map human brain function might be used to measure an individual's propensity for violent or aberrant behaviour — or even, one day, to predict it.

At next week's meeting, the presidential commission will hear from each of the US agencies involved in the BRAIN Initiative — DARPA, the National Institutes of Health and the National Science Foundation — about preliminary scientific plans and anticipated ethical issues. Lisa Lee, the commission's executive director, says that the group plans to discuss broad ethical concerns for human and animal participants in neuroscience research, and also the societal implications of discoveries that could arise from the BRAIN Initiative. Although no specific timeline has been set, the commission typically holds three to four meetings over a period of up to 18 months, culminating in recommendations to the President.

As neuroethicists wade into the issues, they may look to the precedent set by the Human Genome Project's Ethical, Legal and Social Implications (ELSI) research programme, which has provided about US\$300 million in study support over 23 years. The programme raised the profile of genetic privacy issues and laid the foundations for the Genetic Information Nondiscrimination Act of 2008, which prohibits discrimination by employers and health insurers on the basis of genetic information.

Thomas Murray, one of the architects of ELSI and president emeritus of the Hastings Center, a bioethics research institute in Garrison, New York, is among the speakers invited to the commission meeting. He considers the BRAIN Initiative a timely opportunity to develop an ELSI programme for neuroscience. "There will be wonderful questions about human responsibility, human agency," he says. "It's never too soon to begin." ■

1. Ramirez, S. et al. *Science* **341**, 387–391 (2013).
2. Pitman, R. K. et al. *Behav. Neurosci.* **125**, 632–638 (2011).
3. Barak, S. et al. *Nature Neurosci.* **16**, 1111–1117 (2013).
4. Arzi, A. et al. *Nature Neurosci.* **15**, 1460–1465 (2012).



The Skolkovo Institute of Science and Technology will open near Moscow next year (artist's impression).

INNOVATION

Russia pins hopes on science city

But sceptics question prospects for Skolkovo commercial park.

BY DECLAN BUTLER

A Russian revolution in scientific innovation — or a white elephant? Bulldozers are rumbling near Moscow, at work on the Skolkovo Innovation Centre, an ambitious, multibillion-dollar scheme to boost Russia's moribund innovation system.

Scientists have high hopes for the project's first goal: to build a world-class technology university from scratch in a few years. However, they are more sceptical about the prospects of a planned commercial science park at the site, and some have baulked at the high price — a reported US\$300 million — paid to a US institution to jump-start the university. Meanwhile, allegations of corruption at the innovation centre's umbrella body, the Skolkovo Foundation, risk throwing a cloud over the entire enterprise.

Launched by the Russian government in 2010 with 85 billion roubles (US\$2.6 billion) in

state funding until the end of 2014, the Skolkovo supercampus will rise on a 400-hectare site just west of the Moscow ring road. On 1 August, the government announced that it intends to put a further 135.6 billion roubles into the venture by 2020. Focusing on five areas — information technology, nuclear technology, energy efficiency, biomedical innovation, and space and telecommunications — Skolkovo is the boldest of Russia's efforts to spur high-tech innovation and reduce the country's economic dependence on exports of oil, gas and minerals.

The intellectual lynchpin of the enterprise is the Skolkovo Institute of Science and Technology (Skoltech), an elite, English-language, graduate and research university being created

in partnership with the Massachusetts Institute of Technology (MIT) in Cambridge. It aims to foster a new generation — and breed — of

► **NATURE.COM**

For more on attempts to boost Russian science, see:
go.nature.com/u7yyqg



YURI KAZATCHKOFF

top-notch Russian researchers and engineers trained in translational research and entrepreneurship as well as basic science, to fuel an ecosystem of innovation that the country sorely lacks. It plans to recruit 200 full-time faculty, 300 postdocs and 1,200 students by 2020.

Coaxing leading scientists to Russia is a challenge, so Skoltech is offering compensation packages that exceed those available in the West "by a wide margin", says Konstantin Severinov, a molecular biologist and Skoltech associate dean of faculty. Skoltech hopes to have 30–35 faculty members by the end of this year, and then to add 30 or so annually until it reaches its goal.

The most immediate difficulty is practical: Skoltech's first buildings will become available only next year. For now, MIT and other partners are teaching the first students, who started studies last September, and faculty members are working elsewhere. The challenge, says Raj Rajagopalan, a chemical engineer and Skoltech's dean of faculty, is "how to create an academic community when most people are not here".

It was a deliberate choice to start Skoltech's activities in parallel with construction, says Duane Boning, an electrical engineer and MIT's faculty lead on Skoltech. The plan will shave years off the time needed to get the university up and running, he says. It has allowed Skoltech to put its administrative, curriculum and research strategies in place quickly, adds Severinov, who has recruited three postdocs, a programmer and one PhD student for his lab, with more to come soon. But "we really need those buildings here ASAP", he adds.

The terms of MIT's four-year contact with

Skoltech, which ends in 2015, are confidential, but those familiar with the project say that Russia is paying MIT at least \$300 million for its services. Several members of the Skolkovo Foundation's scientific advisory board originally objected to that price tag, say sources. Some argued that more services could be had for less, or from another partner — and others felt that Russian institutions should have been given a greater role in the project.

Alexei Sitnikov, vice-president of institutional and resource development at Skoltech, confirms that some board members expressed reservations, but he defends the deal. "The benefits which we get exceed, by far, the costs that we pay," he says. Several scientists involved say that, although the price can be debated, MIT has been effective in developing the elements needed for the institute to get going: curricula, research programmes, administrative and recruitment structures, and mechanisms for entrepreneurship and innovation. MIT has also run Skoltech's international calls for proposals for research centres of excellence (15 are planned). "Sometimes you can pay a lot for a brand name," adds Rajagopalan. After its contract ends, MIT will shift to a collaborator role and Russian institutions will be more involved in Skoltech, adds Sitnikov.

Building Skoltech "is exactly the right thing to do", says Leonid Levitov, a Russian physicist at MIT who is not directly involved in Skoltech's creation. At first, he doubted that MIT should take a role, given the political complexity, bureaucracy and corruption that exist in Russia.

But he is impressed at how Skoltech, and MIT's contributions to it, are shaping up. Historically, much of the country's most innovative research was linked to government labs, including those run by the military, but most have collapsed since the fall of Soviet communism. These days, there are few graduate opportunities. Many of the best students

leave Russia, and even more switch to non-research careers, he says. Skoltech could be a place for them to train in research at home.

The institute's ultimate impact will depend heavily on reforms elsewhere in Russian science, as well as on the success of the Skolkovo science park. Without it, there will be few meaningful career opportunities for the elite researchers that Skoltech will produce, and the institute could end up as "a completely crazy brain-draining scheme", says Severinov, with Russia training first-rate people who then leave.

The park offers companies incentives including tax breaks and visa help. Twenty-eight ▶



"We really need those buildings here ASAP."

Konstantin Severinov

► major partners, including Siemens, IBM, Intel and Microsoft, have pledged to invest a total of \$500 million, which will be used, in part, to establish corporate research and development centres at Skolkovo. More than 960 start-up companies have signed on to set up shop there, and around 60 venture-capital firms, more than one-third of them from other countries, have pledged a total of \$600 million, says Leonid Gankin, a spokesman for Skolkovo.

But with construction still under way, start-ups will not be able to begin moving to Skolkovo until 2014. Only one corporate centre has been

built: that of Cisco Systems, a networking-equipment firm based in San Jose, California, which opened its Skolkovo facility in June. A report from the European Bank for Reconstruction and Development in London last December warned that although Skolkovo offers benefits, initiatives in which states try to pick winning industries have a chequered history.

Skolkovo's reputation took a hit this year, with allegations of corruption at the Skolkovo Foundation. In February, the Investigative Committee of the Russian Federation, the country's main criminal-investigation authority,

launched enquiries into the alleged embezzlement of 24 million roubles by Skolkovo officials. Investigators raided the foundation's Moscow offices on 18 April.

Levitov and other Russian scientists warn against jumping to conclusions before the charges are substantiated, arguing that in Russia, criminal allegations can often mask political intrigues and power struggles. Still, Rajagopalan worries that some prospective faculty members might ask themselves whether they really want to take the risk of decamping there. "Perceptions count," he says. ■

PHYSICS

Europe sets sights on lasers

Three eastern European countries are gearing up to host powerful light sources.

BY DEVIN POWELL

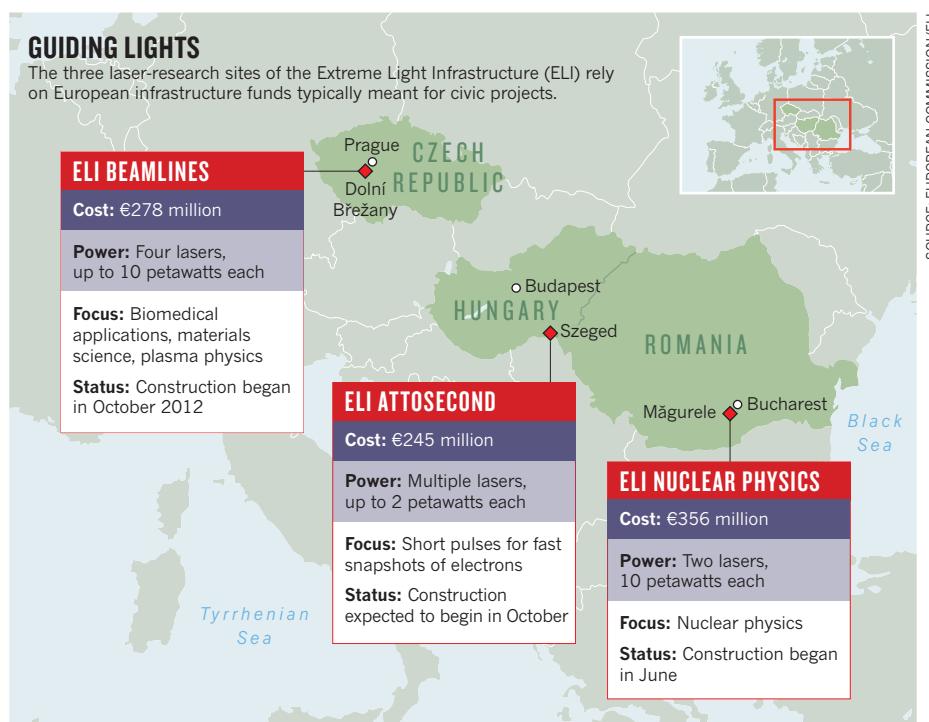
Last month, a large European nation placed an order for a pair of titanium-sapphire lasers. With an output of 10 petawatts, each would be about nine times more powerful than the strongest laser in existence today — capable of exploding the nuclei of heavy elements.

Nearly as special as the lasers is the place where they will be delivered: Măgurele, Romania, a town of 11,000 people that was once home to a Soviet-era research reactor. In June, workers broke ground at the construction site for a new research centre to house the twin lasers.

It would not be possible for the former Eastern bloc nation to afford the project's €356-million (US\$475-million) cost: expenditures on research and development in Romania are among the lowest in Europe. Instead, the European Union (EU) is expected to pick up most of the tab, as it plans to for two other facilities, in Hungary and the Czech Republic. The three projects, along with a fourth still on the drawing board, will make up the Extreme Light Infrastructure (ELI) — an experiment in using EU structural funds to advance research.

"With ELI, we enter a new era, pioneering the use of structural funds for an international research infrastructure," says Wolfgang Sandner, who heads a consortium set up by the three countries and Italy to manage ELI.

Structural funds, which make up more than one-third of the EU budget, are aimed at reducing economic disparities across Europe. They typically pay for local projects such as road repair, construction of power lines or clean-up of industrial waste — not super-powerful lasers. And although any member



state can apply for these infrastructure funds, most of the money goes to the poorest countries.

ELI, by contrast, is an international research institute that will encourage scientists to do laser experiments at higher energies, intensities and speeds. Both France and the United Kingdom wanted to host the project, marked as a priority in a 2006 road map for European research. But structural funds provided the Czech Republic, Romania and Hungary with a way to pay for it. Advocates for using the money in this way touted the potential of ELI to stimulate economic growth. Romania hopes to attract technology companies and establish

Măgurele as a 'laser valley'; the region is also planning to build a smaller laser facility to train its scientists to participate in ELI.

But the structural funds come with strings attached. For starters, the money must be spent by 2015, and meeting that deadline is proving difficult. After Romania was selected for ELI in 2009, it took two years to decide how to shift structural funds away from other projects, delaying the start of construction. And Hungary submitted a formal funding proposal only in May. That request has yet to be approved.

Using structural funds to buy high-end

lasers has also been tough, says Vlastimil Růžička, managing director of the Czech facility, to be built near Prague. That is partly because these funds usually go towards civic projects and products with clear prices and capabilities. Working out the uncertainties involved in developing cutting-edge lasers has required protracted negotiations with laboratories. "Included in the contracts are penalties for not delivering," says Růžička, who has spent nine months working out a deal to buy a laser from the Lawrence Livermore National Laboratory in Livermore, California.

The three countries have asked the European Commission to approve a phased approach that would bring some lasers online in 2015 and others in 2017. Adrian Curaj, general director of Romania's Ministry of Education, Research and Youth in Bucharest, says he is confident that this approach will pan out. "Everyone wants a success story here," he says.

Each branch of ELI will specialize in a different kind of laser science (see 'Guiding lights'). Romania's intense lasers will explore nuclear physics, and the Czech lasers will accelerate particles to make X-ray and ultraviolet light sources, useful in exploring the structure of molecules and solid materials. One laser will fire exceptionally rapid pulses to excite electrons in a plasma. Those electrons then shed that energy, emitting a

laser-like, or coherent, beam of X-rays.

Most sources of coherent X-rays, such as the European X-ray free-electron laser (XFEL), under construction near Hamburg, Germany, use microwaves to accelerate electrons over a long distance — 1.7 kilometres, in the case of the XFEL. By contrast, ELI's compact X-ray laser technology will fit on a tabletop in a small laboratory. Its designers hope that a successful demonstration will encourage wider use of laser-based X-ray sources. The pulses that ELI makes will

be complementary to those at XFEL — less energetic but spread over a broader spectrum — and will be useful for probing different kinds of atoms in a material simultaneously, for instance.

Hungary's lasers, to be housed in a former military barracks in the town of Szeged, will also produce X-rays. But here the goal will not be to make pulses coherent, but to make them as short as possible — each lasting less than one-quadrillionth of a second. At this time scale, the pulses can excite a target atom's electrons, which can then be tracked by the

X-rays. "Each time laser pulses have gotten shorter, a new physical regime has been accessed and we have learned new things," says Todd Ditmire, director of the Texas Petawatt Laser project at the University of Texas at Austin and chief executive of National Energetics, a laser company based in Austin that has plans to bid on building one of Hungary's lasers.

As the ELI trio takes the stage, a fourth ELI project waits in the wings: a facility for relativistic physics, underpinned by an even mightier laser that would reach 200 petawatts of power. This titan's beam would be strong enough to distort empty space — enabling researchers to test the idea that a vacuum can behave like a crystal and bend light. The powerful beam could probe the nature of virtual particles as they pop in and out of the void and search for radiation that is thought to be given off by an object accelerating through a vacuum.

How to make that laser is still an open question — as is where to build it and how to pay for it. The fourth site might go to one of the countries that is already participating. Another option is Russia, which has proposed building a 'super-laser' from 12 smaller ones. "It was, and remains, ELI's strategy to first gain some experience about the feasibility of different technologies at the three existing pillars before narrowing down the options," says Sandner. ■

"Each time laser pulses have gotten shorter, a new physical regime has been accessed and we have learned new things."

India spurns cancer patents

Nation seeks to cap high cost of drugs to treat non-infectious diseases.

BY ERIKA CHECK HAYDEN

Once the scourge of the developing world, infectious diseases such as malaria, tuberculosis and AIDS can now be fought with cheap drugs. But as people in poorer nations live longer and adopt Western habits, non-communicable diseases such as heart disease, diabetes and cancer have become the main killers — and paying for their treatment has become a thorny problem.

India may now be drawing a line in the sand. In the past three weeks, officials there have refused patents on two breast cancer drugs — the latest in a series of decisions to limit patents on pricey brand-name medications. These moves reflect a tension: India now surpasses the United States in terms of annual cancer deaths, and wants to find ways to treat the disease cheaply. But this desire runs counter to the goals of drug makers, who see middle-income nations as central to their growth plans.

The first of the recent rejections occurred on 27 July, when an Indian federal board of patent officials revoked a patent on a slightly modified version of the breast cancer drug lapatinib, sold as Tykerb by London-based pharmaceutical firm GlaxoSmithKline. Then, on 4 August, Swiss drug company Roche reported that

a patent office in the city of Kolkata, a hub of the national patent system, would not grant patents on a version of the company's drug trastuzumab, sold as Herceptin. Indian officials allowed other patents that will protect both drugs from generic competition until 2019. But the rulings will stop the companies from extending their patent protection beyond that date, opening a window for manufacturers of generic drugs to then step in.

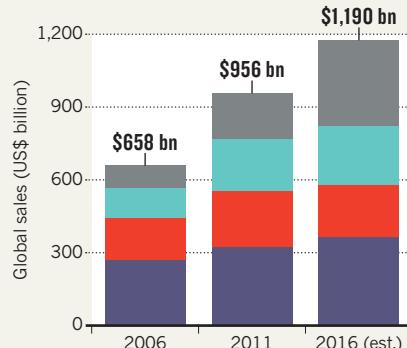
The fight echoes one in the late 1990s and early 2000s over drugs for treating infections such as HIV. That dispute was largely resolved when drug makers allowed developing-world companies to create cheap generic medicines. Today, antiretroviral treatments can be bought for less than US\$100 a year, compared with more than \$10,000 a year in 2000, according to international aid organization Doctors Without Borders (Médecins Sans Frontières), based in Geneva, Switzerland.

"We think the answer is to make the price of drugs really cheap."

DRUG MONEY

By 2016, drug makers could be earning as much from emerging markets, such as China, Brazil and India, as they do from selling in the United States.

■ United States ■ Europe
■ Rest of world ■ Emerging markets



Emerging markets: China, Brazil, India, Russia, Mexico, Turkey, Poland, Venezuela, Argentina, Indonesia, South Africa, Thailand, Romania, Egypt, Ukraine, Pakistan and Vietnam.

But drugs for non-communicable diseases — particularly cancer — will be much trickier to negotiate. "There's no easy compromise that's going to arise around non-communicable diseases the way that we saw around HIV," says Thomas Bollyky, a lawyer with the Council on Foreign Relations in New York who fought for affordable HIV medicines in the 1990s.

In India, a \$15,000 course of trastuzumab can cost more than ten times the average annual wage. And there are no older, off-patent drugs that could serve as an alternative, because none of them target the specific type of breast cancer as well as trastuzumab.

Yet drug makers are reluctant to cut prices in middle-income countries such as India, China and Brazil, which are projected to account for much of the industry's growth in the near future (see 'Drug money'). Although Africa's ability to pay for HIV drugs was never going to be high, some people in middle-income nations can afford expensive medicines. Drug makers do not want to erode that niche market through lower-cost drugs, even if the vast majority of people in need cannot pay, says James Love, director of Knowledge Ecology International, a non-governmental organization in Washington DC that advocates for social justice in access to knowledge.

A number of ideas to skirt the impasse have been floated, but none are simple. Drug makers argue that governments of middle-income nations should broaden insurance

programmes and access to health care. They also argue that drugs will become more affordable as economies grow and people earn more money. By reducing support for intellectual property, India is undermining incentives for drug development and foreign investment that will allow for growth, says Amy Hariati, director and legal policy counsel for life sciences at the US-India Business Council, an industry group based in Washington DC. "The best way for the Indian economy to grow is by rewarding innovation," she says.

Another idea comes from the World Health Organization, which for the past five years has been trying to broker an international treaty that would see member states supporting the development of lower-cost medicines with prizes and research funding rather than patents. "We think the answer is to make the price of drugs really cheap, and to provide funding as a reward for innovation rather than through a monopoly on a drug," says Love.

There is also increasing pressure on drug companies to adopt pricing models that allow people in the same country to be charged different prices for drugs, depending on their ability to pay. Companies, including Roche in the case of trastuzumab, say that they already offer such differential pricing through special access programmes. Still, Roche's own figures show that it sold enough trastuzumab last year to treat only 3,700 Indian breast cancer patients — 15% of those who need it.

The battles may end up being mere skirmishes if India goes further and allows local companies to disregard Roche's trastuzumab patent altogether and manufacture a cheaper generic version, using a 'compulsory licence'. Last year, India issued such a licence on a cancer drug sold by German firm Bayer. And in January, India's ministry of health recommended compulsory licences for trastuzumab and two other cancer drugs.

Indonesia issued compulsory licences for seven drugs in 2012, and China and the Philippines have tweaked their laws to make such licences easier to issue. Prashant Yadav, director of the health-care research initiative at the University of Michigan in Ann Arbor, says that these moves portend an unsettling future. India may be the main battleground today, but the war over cancer-drug access seems likely to bleed beyond its borders unless a compromise is reached. "This requires some kind of diplomacy now," says Yadav. ■

CORRUPTED MEMORY

Elizabeth Loftus has spent decades exposing flaws in eyewitness testimony. Her ideas are gaining fresh traction in the US legal system.

BY MOHEB COSTANDI

In the early hours of 9 September, 1984, a stranger entered Mrs M's California home through an open living-room window. Finding Mrs M asleep, he tried to rape her, but fled when other people in the house awoke. Mrs M described her assailant to the police: he was black, weighing about 170 pounds and 5'7" to 5'9" tall, with small braids and a blue baseball cap.

Officers cruising her neighbourhood spotted someone roughly matching that description standing beside his car a block away from the house. The man, Joseph Pacely, said that his car had broken down and he was looking for someone to jump-start it. But Mrs M identified him as her attacker and he was charged.

At Pacely's trial a few months later, memory researcher Elizabeth Loftus testified on his behalf. She told the jury how memory is fallible; how stress and fear may have impaired Mrs M's ability to identify her assailant, and how people can find it difficult to identify someone of a race other than their own.

Pacely was acquitted. "It's cases like this that mean the most to me," says Loftus, "the ones in which I play a role in bringing justice to an innocent person."

In a career spanning four decades, Loftus, a psychologist at the University of California, Irvine, has done more than any other researcher to document the unreliability of memory in experimental settings. And she has used what she has learned to testify as an expert witness in hundreds of criminal cases — Pacely's was her 101st — informing juries that memories are pliable and that eyewitness accounts are far from perfect recordings of actual events.

Her work has earned her plaudits from her peers, but it has also made her enemies. Critics charge that in her zeal to challenge the veracity of memory, Loftus has harmed victims and aided murderers and rapists. She has been sued

and assaulted, and has even received death threats. "I went to a shooting range to learn how to shoot," she says, noting that she keeps a few used targets in her office as a point of pride.

Now, the 68-year-old scientist's research is starting to bring about lasting changes in the legal system. In July last year, the New Jersey Supreme Court issued a ruling — based largely on her findings — that jurors should be alerted to the imperfect nature of memory and the fallibility of eyewitness testimony as standard procedure. Loftus is working with judges in other states to make such changes more widespread.

"What's going on now in America really is something of a revolution," says Martin Conway, a cognitive psychologist at City University London. Loftus' work, he says, has been "profoundly important" in shaping these changes.

MALLEABLE MEMORIES

Loftus says that her start in psychology was rudderless. As a graduate student in mathematical psychology at Stanford University in California, "I wasn't really captivated", she says. "I'd sit in the back of the seminars, kind of bored, writing letters to my Uncle Joe or hemming skirts, or whatever."

Eventually a social-psychology class piqued her interest and she began to study how word meanings are stored in the brain, and how people recall them. Still, something was missing. "One day I was having lunch with a cousin of mine," she says, "and I told her about our great discovery that people are faster at naming 'a bird that's yellow' than 'a yellow bird'." Her cousin — unimpressed — joked about taxpayers' money being wasted. "That's when I decided I wanted to work on something that had more practical applications."

Loftus was casting about for a meaningful way to study memory

and get funding when a former Stanford engineer working for the US Department of Transportation said that his employer would probably pay for research into car accidents.

Following that lead, Loftus won funding in 1974 for a proposal to study witness accounts of accidents, and she soon published the first of several influential studies revealing the limitations of eyewitness testimony¹. She showed people film clips of car accidents and asked them to estimate the speed of the cars. The wording of the questions, she found, had a profound effect on the estimates. People who were asked, "How fast were the cars going when they smashed into each other?" gave higher estimates on average than those with whom the verb 'hit' was used. And those who were told that the cars had 'contacted' each other gave the lowest estimates.

Those asked about cars smashing into one another were more than twice as likely as others to report seeing broken glass when asked about the accident a week later, even though there was none in the video. "I realized that these questions were conveying information," says Loftus. "I began to think of it as a process of memory contamination, and we eventually called it the misinformation effect."

She went on to publish several other studies^{2–4} showing how memories can be contorted, and that the ability of eyewitnesses to identify suspects from photographs can be unreliable. Any description they might hear has the potential to influence who or what they think they saw.

Loftus was eager to translate these findings to the real world, and began consulting on legal cases to get "close up and personal" with witnesses. Her first case — of a woman accused of killing her abusive boyfriend — hinged on whether the woman had acted in self-defence or had committed premeditated murder. Eyewitnesses could not agree on how much time had elapsed between when the defendant had picked up the gun and when she had fired it: some said

NATURE.COM

For a slideshow and podcast about Elizabeth Loftus, see: go.nature.com/Bcguiw

it was seconds, others said minutes. Loftus cast doubt on the memory of the witnesses, and the woman was acquitted.

Loftus described the case, together with her research, in a 1974 article for *Psychology Today* magazine⁵. "Once that article came out, I started getting calls from all over the place," she says. "From lawyers wanting me to work on their cases, and legal professionals wanting me to lecture at their meetings."

Some of her trials have been high-profile — including that of the serial killers known as the Hillside Stranglers and the 1992 trial of the police officers indicted for beating construction worker Rodney King. She even consulted on a case involving a young law student named Ted Bundy, who was accused of kidnapping a woman in 1974. Bundy was convicted, only to escape. Recaptured in 1978, he eventually admitted to killing 30 people.

The possibility of aiding guilty people does not faze Loftus. "I haven't had a situation where someone was acquitted because of my testimony and then went on to commit some awful crime," she says. "I would feel horrible if that happened, but I'm only one small piece of a court case." She is often compensated for her expert-witness work, earning up to US\$500 per hour, she says.

Nita Farahany, a bioethicist at Duke University School of Law in Durham, North Carolina, says that Loftus's activism in the court is not unique, and that testifying on behalf of unpopular defendants is important. "It shows that she has tried to be truly impartial, and that her goal is to try to provide an accurate understanding of the science, no matter who is involved."

Still, Loftus has drawn the line at some defendants, such as John Demjanjuk, who in 1988 stood accused in Israel of being 'Ivan the Terrible', a guard who operated gas chambers at the concentration camp Treblinka in Poland during the Second World War. Loftus, herself Jewish, declined to testify because she worried that it would upset family and friends.

The case led some to accuse her of double standards. But those criticisms were mild compared with the reactions that she would soon trigger in her most controversial legal work.

DIGGING UP THE PAST

In 1990, Loftus got a call from a California attorney defending George Franklin, whose daughter claimed that during therapy, she had recovered decades-old memories of him murdering her friend, Susan Nason. Loftus decided to consult for the defence team. "I thought it was pretty fishy and started looking into the literature," she says. She found little convincing research to support the idea that traumatic memories could be repressed for years.

Franklin was convicted despite her testimony. He spent five years in prison before an appeals court reviewed and then overturned his conviction amid doubts over his daughter's statements.

The courts went on to see a surge in cases



based on recovered childhood memories, fuelled in part by popular books and high-profile accusations. Loftus began to wonder whether it was possible to fabricate complex, believable memories. "I wanted to see if we could implant a rich memory of an entirely made-up event," she says. An idea eventually came to her as she drove past a shopping mall.

Working with a student, Jacqueline Pickrell, Loftus recruited 24 people and, with the cooperation of family members, presented them with four detailed accounts of events from their childhood. Three of the incidents had actually taken place, but the fourth — a dramatic account of being lost in a mall — was entirely concocted by Loftus and corroborated by the participants' relatives. One-quarter of the participants claimed to remember the false event⁶.

BATTLE GROUND

Loftus became convinced that well-meaning psychotherapists could inadvertently implant false memories into patients' minds, and her subsequent testimonies led to a row between therapists who believed their patients were recovering lost memories and researchers who thought something else was afoot. To try to settle these 'memory wars', the American Psychological Association (APA) commissioned an expert report about the subject, to be written by three memory researchers, including Loftus, and three clinical psychologists.

The groups could not agree, and each ended up writing a separate report. "It was very polarizing," says Stephen Ceci, a developmental psychologist at Cornell University in Ithaca, New York, who worked with Loftus on one of the reports.

There are ways in which traumatic memories of real events can be recalled after being buried for years, he adds, but without hard evidence, it is impossible to distinguish false memories from real ones in court. It is, therefore, possible that some claims of childhood abuse go unvindicated because of Loftus' testimony, and this is the cause of much of the hostility towards her.

Ross Cheit, a political scientist at Brown University in Providence, Rhode Island, started the Recovered Memory Project in 1995 to document and respond to what he says has been a one-sided debate. There are now more than 100 corroborated cases of recovered memory on his website (<http://blogs.brown.edu/recovered-memory>), he says, including some on which Loftus had consulted.

"Loftus is often on the losing side, and she's sometimes wrong in a spectacular way," Cheit says. Her testimonies, he adds, can be psychologically damaging for the victims. "If you're telling someone you think their memories are false, when they have corroborating evidence that they were abused, that's corrosive."

Loftus does not believe that Cheit's site corroborates recovered memories. "He might have some cases of people who didn't think about

their abuse for some time and were reminded of it, but as for actual repression, no," she says. "I cringe at the idea of hurting genuine victims, but when an innocent person is accused, we have a whole new set of victims, and I'm more horrified by an innocent person getting convicted than by a guilty person being acquitted."

But her testimonies and investigations into recovered memories have strained her professional relationships. Towards the end of 1995, two women filed formal complaints against Loftus with the APA. Lynn Crooks and Jennifer Hoult had won civil suits in cases involving recovered memories of childhood sexual abuse, and both claimed that Loftus had distorted the facts of their cases in articles and interviews. Loftus resigned from the APA and critics speculated that she had caught wind of the complaints and left before a formal investigation could take place. But Loftus chalks her resignation up to political disagreements, saying she knew nothing of the complaints at the time.

In 1997, Loftus and several colleagues began to dig into a published case study describing an anonymous subject, 'Jane Doe', who had apparently recovered a repressed memory of childhood abuse⁷. They found information that cast doubt on her account, but before they could publish, Doe contacted the University of Washington in Seattle, where Loftus was working, and accused the team of breaching her privacy.

The university confiscated Loftus's files, put her under investigation for nearly two years and prevented her from publishing. She was eventually

"I BEGAN TO THINK OF IT AS A PROCESS OF MEMORY CONTAMINATION."

ally cleared, and published the work⁸ in 2002. The next year, however, Doe sued Loftus and her collaborators for fraud, invasion of privacy, defamation and causing emotional distress.

It was at around that time that Loftus moved to the University of California, Irvine. The Jane Doe case was eventually settled in 2007, when the Supreme Court of California dismissed all but one of the charges and Loftus agreed to pay a nuisance settlement of \$7,500. "It was such a stressful time, but I can't really say it was detrimental overall," says Loftus.

Her work has now moved from trying to affect single cases to pushing for broader changes in the legal system. Loftus has been working with Pennsylvania trial judge Jeannine Turgeon to compile a set of guidelines similar to those instituted in New Jersey last year. They instruct jurors that memory "is not like a video recording" and ask them to consider the many factors that can alter memories, such as the presence of a weapon, which can draw attention

away from the perpetrator's face.

"This has the potential to be really important," says Farahany. "Using cutting-edge research to undercut the idea that memory is as stable and precise as people believe it to be can really help us get to a place where we have better truth-seeking in criminal cases," she says.

Loftus wants to go further. Almost every stage of the legal process — from the identification and questioning of suspects to cross-examination of eyewitnesses in the courtroom — is prone to error. In a line-up, for example, police officers can influence identification, but this can be avoided if someone who does not know the identity of the prime suspect conducts the line-up (see *Nature* **453**, 442–444; 2008). "I'd like to see this kind of thing being implemented, and to keep educating people about the workings of memory," says Loftus.

MIND CONTROL

Meanwhile, her research has shifted into new controversial waters. Taking on board the lesson that memories can be manufactured, she has been investigating the possibility of using those memories to modify behaviour^{9,10}. "We've shown that you can plant a memory of getting sick eating particular foods as a child," she says, "and we can get people thinking they got sick drinking vodka, so they don't want to drink as much of it later on."

There is no evidence that any of this will successfully transfer from the lab to the real world. Even if it does, it would violate therapists' code of conduct, and could have unforeseen consequences.

"Lying to children is a slippery slope that makes me uncomfortable," says Judy Illes, a neuroethicist at the University of British Columbia in Vancouver, Canada. "Can't we alter their behaviour in a positive way, instead of using subterfuge?" But Loftus dismisses the concerns, suggesting that even if therapists cannot do it, parents might want to. "Parents lie to their kids all the time, about Santa Claus and the tooth fairy. Would you rather have an unhealthy kid, or one with a few false memories?" ■

Moheb Costandi is a freelance science writer in London.

1. Loftus, E. F. & Palmer, J. C. *J. Verb. Learn. Behav.* **13**, 585–589 (1974).
2. Loftus, E. F. *Cognitive Psychol.* **7**, 560–572 (1975).
3. Powers, P. A., Andriks, J. L. & Loftus, E. F. *J. Appl. Psychol.* **64**, 339–347 (1979).
4. Loftus, E. F. & Greene, E. *Law Hum. Behav.* **4**, 323–334 (1980).
5. Loftus, E. F. *Psychol. Today* **8**, 116–119 (1974).
6. Loftus, E. F. & Pickrell, J. E. *Psychiat. Ann.* **25**, 720–725 (1995).
7. Corwin, D. L. & Olafson, E. *Child Maltreat.* **2**, 91–112 (1997).
8. Loftus, E. F. & Guyer, M. J. *Skeptical Inquirer* **26**, 3 (2002). Available at go.nature.com/pr8xp.
9. Bernstein, D. M., Morris, E. K., & Loftus, E. F. *Soc. Cognition* **23**, 11–34 (2005).
10. Clifasefi, S. L., Bernstein, D. M., Mantonakis, A. & Loftus, E. F. *Acta Psychol.* **143**, 14–19 (2013).

COMMENT

HISTORY The polarization of US environmental policy over the past 50 years p.273



PHOTOGRAPHY Ex-medic chronicles stages of life with daguerreotypes p.274

ENVIRONMENT Better palm-oil yield could boost, not slash, plantation area p.276

CONSERVATION Search diplomats' bags to stop them smuggling ivory p.276

JACK DYKINGA/NATUREPL.COM



The unique Florida Everglades deserve protection, despite being highly vulnerable to climate change.

Climate change must not blow conservation off course

Reconfiguring protection priorities around global warming could be of limited use or even harmful, say **Morgan W. Tingley, Lyndon D. Estes and David S. Wilcove**.

In March this year, the United States released its national strategy for conservation planning in the face of climate change. The first goal of the National Fish, Wildlife and Plants Climate Adaptation Strategy is to “conserve habitat to support healthy fish, wildlife, and plant populations and ecosystem functions in a changing climate”. It warns that by prioritizing certain species and areas over others, there will be winners and losers. Elsewhere in the conservation community, there is increasing acceptance of abandoning protection

for many of the species, populations and ecosystems that are most at risk of extinction from climate change¹.

Although climate change seems likely to wreak havoc on biodiversity, as is already happening in some places, its precise effects are difficult to predict. In the meantime, numerous threats that are better understood and more immediate — notably, the direct destruction of habitats — continue to drive species towards extinction.

The best conservation response to global warming is not to beat an orderly retreat

while saving the strongest, but to consider climate change as one of a suite of maladies, all of which must be addressed to protect biodiversity. In some cases climate change may be the most urgent threat; in most cases it is not².

MOUNTAIN BIAS

In the growing literature on conservation and climate change, the golden word is ‘resilience’. Conservationists hope to take actions that help organisms and ecosystems to survive in a warmer and more volatile world. ▶

► Because the resilience of individual species to climate change is difficult to predict, researchers have suggested prioritizing regions that are expected to be climatically stable for at least the next 100 years³. In the past five years, scientists have undertaken global and regional mapping of climate-change velocity to identify these 'refugia'⁴.

Climate-change velocity indicates how fast (for instance, in kilometres per year) and in what direction temperatures are shifting across a landscape. Velocity calculations suggest that in a changing climate, topographically diverse areas such as mountain chains will be more climatically stable than less-complex terrain such as flatlands. In mountainous areas, a wide range of temperatures occurs in a relatively small area. Thus, the distance an organism must migrate to remain at a constant temperature in a changing climate is much shorter in rugged country than in flat terrain.

Our concern is that using metrics such as climate velocity to guide conservation prioritization may well prove ineffective, or even harmful, given the enormous uncertainty over how species will actually respond to climate change. Meta-reviews assessing the impact of global warming on hundreds of species around the world can give the impression that organisms are uniformly marching uphill or to higher latitudes in step with changes in mean temperature⁵. But averaging behaviour across diverse species to find overall trends may be of little use in predicting what will actually happen in any one location.

In fact, the closer scientists look at species' ongoing responses to our warming planet, the more surprises they uncover⁶. For instance, the ranges of only 51% of the bird species in California's Sierra Nevada mountains moved to higher elevations between 1911 and 2009, despite a 1–2 °C mean temperature increase over that period⁶. The remaining species either shifted to lower elevations or did not move.

Basing conservation priorities on an area's expected resilience to global warming, and thus biasing protection towards mountainous landscapes, risks potentially losing the diversity of flatlands. On the whole, flatter ecosystems tend to be in greater need of protection than rougher terrain; national parks, for example, more commonly exist in mountainous areas because such regions are less suited to cultivation and pasturing⁷.

Prioritizing conservation according to an area's expected resilience at a national or continental scale could mean favouring the Rocky Mountains, the high Andes or Mount Kenya over Florida's fabled Everglades, the *cerrado* of Brazil or Africa's savannahs. Although topographically diverse areas may contain larger numbers of species per unit area than flatlands, many areas of flat terrain are just as important for endemic species. The



LUIZ CLAUDIO MARIGO/NATUREPL.COM

Brazil's cerrado is a tropical savannah that has one of the richest endemic floras in the world.

Brazilian *cerrado*, for example, comprises more than 2 million square kilometres and has one of the richest endemic floras in the world. Yet only around 2% of it is currently protected, and around half the total area has already been cleared for agriculture^{8,9}.

THREAT BY THREAT

To move forward, conservationists should consider all threats to biodiversity when deciding which species, habitats or areas to protect, and should not assume that some threats are more important than others. For example, for any conservation area, each potential threat should be evaluated and weighted by the risk it poses, with full consideration of its severity, reversibility, immediacy, pace and certainty.

Finding ways to combine these factors into a meaningful metric remains a formidable challenge. Various methods exist to help conservationists to factor climate-change uncertainties into their priority-setting, but as yet there is no consensus on how the future threat of climate change should be compared to ongoing and more certain threats, such as land-use change. Creating a consistent approach for prioritization should be a primary goal. The International Union for Conservation of Nature Red List of Threatened Species could be a good model, because it simultaneously evaluates the risk of multiple threats, from pollution to invasive species.

In the meantime, there are several ways for conservation planners to guard against overcompensating for climate threats. When it is possible to assess the uncertainty associated with a particular threat — a 20% reduction in rainfall by 2050, say — this uncertainty can be factored into prioritization schemes. In other situations, models predicting an optimal prioritization scheme can be run several times, with and without different threats incorporated. Basing prioritization

on results obtained from a range of models, each with different parameters, is more likely to result in smart conservation.

Given the sobering picture that has long been emerging from studies and reports such as those produced by the Intergovernmental Panel on Climate Change, it is not surprising that conservationists feel the need to alter course to deal with the threat of climate change. Indeed, some may argue that there is little point in trying to save species from immediate threats when their long-term survival is in question. But if conservation planners don't use resources efficiently to maximize the protection of biodiversity and habitats from today's threats, there may be little left to protect from the effects of global warming 50 to 100 years from now. ■

Morgan W. Tingley and Lyndon D. Estes are at the Woodrow Wilson School, Princeton University, Princeton, New Jersey 08544, USA. **David S. Wilcove** is at the Woodrow Wilson School and at the Department of Ecology and Evolutionary Biology at Princeton University.
e-mails: mtingley@princeton.edu; lestes@princeton.edu; dwilcove@princeton.edu

1. Hagerman, S., Dowlatabadi, H., Satterfield, T. & McDaniels, T. *Global Environ. Change* **20**, 192–207 (2010).
2. Jetz, W., Wilcove, D. S. & Dobson, A. P. *PLoS Biol.* **5**, e157 (2007).
3. Game, E. T., Lipset-Moore, G., Saxon, E., Peterson, N. & Sheppard, S. *Global Change Biol.* **17**, 3150–3160 (2011).
4. Loarie, S. R. *et al.* *Nature* **462**, 1052–1055 (2009).
5. Chen, I.-C., Hill, J. K., Ohlemüller, R., Roy, D. B. & Thomas, C. D. *Science* **333**, 1024–1026 (2011).
6. Tingley, M. W., Koo, M. S., Moritz, C., Rush, A. C. & Beissinger, S. R. *Global Change Biol.* **18**, 3279–3290 (2012).
7. Joppa, L. N. & Pfaff, A. *PLoS ONE* **4**, e8273 (2009).
8. Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, G. A. B. & Kent, J. *Nature* **403**, 853–858 (2000).
9. Silva, J. F., Farinas, M. R., Felfili, J. M. & Klink, C. A. *J. Biogeogr.* **33**, 536–548 (2006).

The Bet examines contrasting views of the future, as in this 1980s *New Yorker* cartoon.

FRANK MODELL/NEW YORKER CARTOON COLLECTION/CARTOONBANK.COM



ENVIRONMENTAL POLICY

The biggest wager

Jon Christensen unpacks the fraught story of a biologist, an economist, and the polarization of US environmental policy.

In 1980, economist Julian Simon challenged biologist Paul Ehrlich to bet on the future price of a basket of raw materials then worth US\$1,000. Ehrlich and two of his colleagues chose five metals crucial to the economy at the time: chromium, copper, nickel, tin and tungsten. Ehrlich bet that prices would rise because of increasing scarcity and demand. Simon bet that they would not. The loser would pay the difference in price after a decade. Simon was at greater risk: prices could have risen indefinitely. But Ehrlich lost: in October 1990, he sent Simon a check for \$576.07. End of story, right?

Not so fast, argues historian Paul Sabin, who in *The Bet* attempts to use their wager to narrate parallel biographies of Simon and Ehrlich, as well as a US political environmental history of the past half-century. The result is a revealing tale. We see the evolution of Simon's 'cornucopian' view of ongoing population growth supported by human ingenuity; his open conflict with Ehrlich's steadfast 'neo-Malthusian' argument that there are real ecological limits to growth; and how both standpoints have influenced US environmental politics. So extreme were their stances that they could be viewed as fomenting a "bitter contest

over the future", Sabin writes.

For instance, Ehrlich famously predicted in his 1968 book *The Population Bomb* (Ballantine Books) that hundreds of millions would starve to death in the coming years. A year later he said, "By the year 2000 the United Kingdom will be simply a small group of impoverished islands, inhabited by some 70 million hungry people ... If I were a gambler, I would take even money that England will not exist in the year 2000." By contrast, Simon viewed the population explosion as "a triumph for mankind". Humanity, Simon believed, was "the ultimate resource". He held that human enterprise would continue to solve

The Bet: Paul Ehrlich, Julian Simon, and Our Gamble over Earth's Future
PAUL SABIN
Yale University Press:
2013.

emerging problems and improve life on Earth, and as resource supplies diminished, prices would rise, driving discovery of more reserves or the creation of substitutes.

In *The Bet*, Sabin traces these competing ideas through the energy crisis of the late 1970s, attempts at US immigration reform in the 1980s and the stand-off over climate change in the 1990s. He sees echoes of the conflict embodied in Simon and Ehrlich's wager in Jimmy Carter's loss to Ronald Reagan in the 1980 presidential election, the environmental movement's ongoing lack of a unifying leader, and even the paralysing political divide over climate change in the United States today. Ultimately, the bet is used to explain the whole messy evolution of US environmental politics from the early 1970s, when Republican Richard Nixon was an environmental champion, to today, when Republican environmentalists are an endangered species.

Sabin recognizes the complexities of the larger political story he tries to tell. He was a participant on its edges as director of the



Julian Simon.



Paul Ehrlich.

non-profit Environmental Leadership Program before joining the history faculty at Yale University in New Haven, Connecticut. And he acknowledges that there are other explanations for the divide in thinking on US environmental policy: the shift of the Democratic and Republican parties to more ideologically consistent blocs, business-led backlashes against environmental regulation, and the success of conservative think tanks in staking out positions in public-policy debates on these issues. But he

"Ultimately, the bet is used to explain the whole messy evolution of US environmental politics."

wants to convince the reader that the clash is not played out only in the corridors of power. The wager, he asserts, "stands for much, much more — our collective gamble on the future of humanity and the planet".

Sabin uses the bet as a synecdoche — a narrative device in which a part stands for a whole, in this case environmental politics. In the end, this is simplistic and blurs cause and effect, explanation and interpretation. Sabin bemoans the polarization that is the very structure of a wager, as if it were an important cause of the larger divide. But bets can be useful — they can clarify what is at stake and, by doing so, help us to frame thinking about the future of the planet.

On a personal level, this is a sad tale of two very smart men who talked past each other for years, until in one telling moment they put their convictions on the line. After their wager was decided, they descended into ad hominem attacks. It is depressing to think that Sabin might be partly justified in making this a metaphor for the environmental politics of our age.

On the other hand, there is the story of John Holdren, a secondary character in *The Bet*. A physicist who co-authored several books with Ehrlich and joined his wager against Simon, Holdren is now US President Barack Obama's senior science adviser. How does Holdren's boss bet on the future? "I can't predict what will happen over the next 40 years," Obama once replied when asked to wager his own bet. But, he continued, "I am — and always will be — full of hope about what the future holds". ■

Jon Christensen is an adjunct assistant professor in the Institute of the Environment and Sustainability and the Department of History at the University of California, Los Angeles.
e-mail: jonchristensen@ioes.ucla.edu



Perpetual Disunity, a photo by Mark Kessell from his *Perfect Specimens* series.

Q&A Mark Kessell

Life-cycle imager

Medic-turned-artist Mark Kessell creates prints evoking evolution and human development using the early photographic form of the daguerreotype. As his latest show opens in New York, he talks about shooting portraits of primates, foetuses, the nearly dead and the newly born.



What is *Perfect Specimens*?

It's a photographic life cycle of *Homo sapiens*. I tried to make images that anyone could recognize as distinctly human, showing conception, fetal development, birth, senescence and death. I want the photographs to raise questions about what it means, from a biological perspective, to be human.

Can you tell me about the fetal specimens?

There is a bizarre little human skeleton from around 1890, with enlarged eye sockets and no brain. The way it is mounted in a bottle, grinning upwards and perched on a little spike, seems to mock the tragedy for baby and mother. There is also a perfectly normal four-month-old fetus that is uncurled and standing upright and looks like a little alien. You can see the blood vessels under its translucent skin; the top of the skull

has not yet fused. In another bottle there are tiny identical triplets, spontaneously aborted. These little guys once shared the same blood, but now they are forever alone.

How do you capture birth and death?

For birth, you find a brave woman with a generous partner, and as the baby comes out you stand between her legs with a camera. It's unforgettable and inspiring, but also disturbing as that huge head emerges through such a small aperture. For death, I get permission from relatives to document the final moments of their loved one. One photo shows the lips of a woman who has spoken her last words. Many avoid the subject, but death is a biological process, so why pretend?

What part does your medical training play?

I didn't take pictures until I was 39, when, on a road trip around Australia, my girlfriend handed me a disposable plastic camera. Within a year, I'd given up medicine and was attending art school in Manhattan. At first, I imagined shooting remote landscapes for *National Geographic*, but then I thought,

Mark Kessell:
Perfect Specimens
Last Rites Gallery,
Manhattan, New
York. 17 August–21
September.

'Why fight my past?' I thought back to Egbert, the cadaver I dissected in medical school. He was sliced lengthwise so you could see the brain inside the skull. How vulnerable this source of our humanness seemed. And when we had to dissect a human arm, I was awestruck by the perfection of the hand, built with a sophistication far beyond technological ingenuity.

Are you questioning whether we are unique?
Most behaviours that were considered uniquely human until Darwin's time exist in other species. Crows have a sense of play, form social alliances and consider the future in their behaviour. In my series *Unmet Friends*, I explore our capacity to understand other creatures by showing primates, birds and reptiles in poses that appear to express human emotions. In *Specimen Box*, a thousand photos of animal specimens are pinned to the walls, floor and ceiling of a small room. Everything in the room is dead except for the human observers, whose curious faces are captured by hidden cameras.

What other subjects have you covered?
For *Florilegium*, I photographed surgical instruments, such as facelift forceps and urethral dilators, to make them look alive. These are the most menacing of my images. (One was used as the poster for a 2005 horror movie, *Hostel*, which I am not brave enough to see.) I'm not sure I realized how disturbing they were while shooting. These instruments are designed to rend human flesh and alter the body that largely defines our identity.

Does your work overturn assumptions?
I want people to examine their assumptions the way a scientist does. So in every series there is something that isn't what it seems. Many viewers perceive the instruments in *Florilegium* as botanical specimens. In *Unmet Friends*, animal faces may be misconstrued as human and vice versa. In *Perfect Specimens*, it can be hard to tell dead bodies from living ones. I want people to ask, 'Why is the artist showing me this?'

Why use the daguerreotype process?
It's an incredibly beautiful process, but very hard to control. The result depends on how the silver plate is polished, and on the intensity of the chemicals and light. Areas of overexposure turn an exquisite shade of blue. As a doctor, I was taught to control and standardize. But a daguerreotype can't be precisely replicated. Because it is a pure silver plate, it is also a mirror. You'll see your own face reflected. I'll see mine. ■

INTERVIEW BY JASCHA HOFFMAN

'Why fight my past?' I thought back to Egbert, the cadaver I dissected in medical school. He was sliced lengthwise so you could see the brain inside the skull. How vulnerable this source of our humanness seemed. And when we had to dissect a human arm, I was awestruck by the perfection of the hand, built with a sophistication far beyond technological ingenuity.

Books in brief



What Makes a Hero? The Surprising Science of Selflessness

Elizabeth Svoboda CURRENT (2013)

Science writer Elizabeth Svoboda examines the super-altruism we call heroism, drawing on current research to unravel its biological and environmental roots. Bioengineer Karl Deisseroth, for instance, uses light-sensitive algal proteins to pinpoint the brain circuits that control nurturing impulses in mice. And psychologist Philip Zimbardo isolates efficient problem-solving in tight spots as central to heroic acts. Weaving in gripping case studies — such as Christoph von Toggenburg's lifelong fund-raising for the vulnerable — Svoboda concludes that heroes are made, not born.



How Things Shape the Mind: A Theory of Material Engagement

Lambros Malafouris MIT PRESS (2013)

Is the mind imprisoned in the brain? In this mix of neuroscience and philosophy, Lambros Malafouris suggests that mind and materiality are allied in ways that defy reductive world views. He argues that the act of making objects such as knapped tools in prehistory shaped neural processes; and furthermore, that this is a two-way street in which "human intelligence 'spreads out' beyond the skin into culture". Engrossing, from his analysis of how the interplay of hands, neurons, clay, a moving wheel and social context result in a thrown pot, to the links between a blind man's stick and brain plasticity.



Future Bright: A Transforming Vision of Human Intelligence

Michael E. Martinez OXFORD UNIVERSITY PRESS (2013)

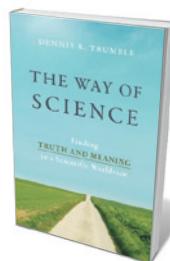
Intelligence can be learned, argues educationalist Michael Martinez. Noting that tackling planetary crises hangs, in part, on the focused work of many agile minds, Martinez shows how such a revolution is possible. Beginning with the deep structure of cognitive ability, such as 'fluid' intelligence — the capacity to adapt to the unfamiliar — he paces the landscape of intellect. His journey reveals tools for change, including the rich variety of intelligence, from social to creative; the brain's extraordinary adaptability; and the cultivation of supportive behaviours such as the urge to excel.



The Neanderthals Rediscovered: How Modern Science is Rewriting Their Story

Dimitra Papagianni and Michael A. Morse THAMES & HUDSON (2013)

Recent findings in geology, genetics and archaeology have radically changed our understanding of Neanderthals. In the first complete chronological narrative of the species from emergence to extinction (perhaps 250,000 to 25,000 years ago), archaeologist Dimitra Papagianni and science historian Michael Morse have shaped a gem. Our big-brained relatives buried their dead, cared for the disabled, hunted creatively and ate grains, wild herbs and even dolphins. A beautifully synthesized portrait of a powerful people.



The Way of Science: Finding Truth and Meaning in a Scientific Worldview

Dennis R. Trumble PROMETHEUS (2013)

The popular perception of science hinges largely on medical and technological advances. Biomedical engineer Dennis Trumble calls for more: a widespread reconnection to science as a way of knowing ourselves and the world. Rationality and critical thinking, he asserts, are moral pathways. Quoting Charles Darwin's "there is grandeur in this view of life", Trumble aims — without histrionics — to urge the religious towards science as a source of meaning. Barbara Kiser

Correspondence

HFEA responds on germline therapy

As chief executive of the UK Human Fertilisation and Embryology Authority (HFEA), I take issue with Marcy Darnovsky's criticisms of our public consultation on mitochondria replacement in reproductive therapy (*Nature* 499, 127; 2013).

Darnovsky queries the statistic used by the HFEA: "one in 200 children is born each year with a form of mitochondrial disease." This comes from a study of mitochondrial mutations in the general population (H. R. Elliot *et al.* *Am. J. Hum. Genet.* 83, 254–260; 2008) and includes children with mild or no symptoms. Other UK organizations use the same figure, notably the National Health Service (see go.nature.com/dcocom2) and the Nuffield Council on Bioethics (see go.nature.com/s6z8gk).

Darnovsky also questions parts of the HFEA's methodology and conclusions. The consultation was designed with external experts, and enabled public participation in various 'strands' that used qualitative and quantitative analysis. Our aim was to go beyond 'yes' or 'no' responses. In all but one of five strands, most people supported mitochondria-replacement techniques. Our publicly open questionnaire drew more opposition from its self-selected respondents, with some arguing that any *in vitro* manipulation of embryos is inappropriate or unethical.

All the consultation material is available at our website (www.hfea.gov.uk/7784.html). I suggest that people judge for themselves. **Peter Thompson** HFEA, London. peter.thompson@hfea.gov.uk

Ireland's research output set to slide

In the first global Nature Publishing Index (see go.nature.com/vbqdr), Ireland is ranked 20th in the Global Top 100 for

2012 and is identified as one of five countries to watch, on the basis of a marked increase in the number of its publications in Nature Publishing Group research journals. Watchers could be disappointed when the next index is released.

Ireland's research output is under threat. Research career frameworks are being formulated and implemented by several major Irish universities. These effectively limit researchers' careers to six years through the enforcement of an 'up or out' policy, whereby they must either secure increased funding to rise through the ranks, or leave the university.

There is now a real risk that Ireland's talented researchers will be forced out as they reach the end of their allotted research time. They are left with the choice of going into industry, moving abroad to continue their research, or dropping out of science.

Mark Jessopp, Ruairí O'Reilly, Gordon Dalton University College Cork, Ireland. m.jessopp@ucc.ie

New electricity grids must develop in sync

Future US and European electricity systems will depend on the harmonious evolution of low- and medium-voltage 'smart' grids with high-voltage 'super' grids (see M. Amin *Nature* 499, 145–147; 2013). Without care, these two developments could undermine, rather than reinforce, one another.

Both smart and super grids involve three layers (see go.nature.com/7flk7s): physical (power flows), cyber (information technology) and socio-economic (stakeholders). There are important issues in the physical layer, but the greatest tensions will emerge in the socio-economic layer as self-interested stakeholders advance local or continental interests.

Smart grids relate to local consumers becoming electricity

producers themselves. This could diminish the need for high-voltage transmission grids as local power starts to serve nearby users.

Super grids will help to integrate massive renewable sources and to ensure reliability. Their development will be led by transmission companies at a time when those businesses must adapt to emerging smart grids.

If smart and super approaches are not aligned, the result will be a fraileer electricity system. It is not just technologies that need to match up: policy regulations and market structures must do so too.

Smartness should not just be an issue for better local grids, but for the electricity system as a whole.

Ettore Bompard, Marcelo Masera European Commission Joint Research Centre, Institute for Energy and Transport, Petten, the Netherlands.

William J. Nuttall The Open University, Milton Keynes, UK. william.nuttall@open.ac.uk

palms, and almost half of these fires occurred in areas with a moratorium for new licences (see go.nature.com/doiwwf).

Finn Danielsen Nordic Agency for Development and Ecology, Copenhagen, Denmark. fd@nordeco.dk

Faizal Parish Global Environment Centre, Selangor, Malaysia.

Seize diplomats smuggling ivory

One route for the smuggling of wildlife parts, such as rhino horn and elephant ivory, could be closed without the need for new international conventions or treaties (B. I. Kaaria and N. L. Muchiri *Proc. 9th Int. Conf. Environ. Compliance Enforc.* 204–208; 2011). Rogue diplomats must be stopped from exploiting their consular immunity to flout customs regulations and support this illicit trade.

The scale of the problem is unquantifiable, but anecdotal evidence indicates that it is widespread. Diplomatic bags containing horn were reported to be evading customs checks as long ago as the 1980s (E. B. Martin *Pachyderm* 21, 28–34; 1996). In 2011, a South Korean diplomat was caught smuggling ivory (see go.nature.com/fuo43s).

Countries could adopt the UK practice of giving full immunity only to diplomats who are accredited to that country; others who pass through British territory are subject to regular checks. Because there are few direct flights between African and Asian capitals, diplomats would run the risk of also being checked at European airports when changing flights.

Apprehending diplomats carrying illegal materials would generate negative publicity, forcing countries to punish errant civil servants and reduce the problem. Immunity should not mean impunity.

Kelvin S.-H. Peh University of Southampton, UK. kelvin.peh@gmail.com

Composite for smarter windows

Glass has been prepared that selectively absorbs visible and near-infrared light when an electrochemical voltage is applied. This opens the way to 'smart' windows that block heat on demand, with or without optical transparency. **SEE LETTER P.323**

BRIAN A. KORGEL

Residential and commercial buildings account for about 40% of energy use and about 30% of energy-related carbon emissions in the United States¹. To decrease this energy demand, materials are needed that help to regulate the heating and lighting requirements of buildings by responding to environmental changes. In particular, electrochromic window materials, which change colour and/or transparency when subjected to an electric field, could significantly reduce energy consumption in buildings². On page 323 of this issue, Llordés *et al.*³ report a great advance in the development of such materials. They have made a composite in which nanometre-scale crystals of indium tin oxide are embedded in a niobium oxide glass, with high control of nanocrystal loading and dispersion. The electrochromic performance of the composite is much better than expected from a simple sum of the optical absorption of its two separate components, because of both the nanostructure of the material and synergistic interactions that occur at the interface between the components.

Inorganic nanocrystals are typically synthesized chemically with organic capping groups attached to aid the crystals' dispersibility in solvents and to prevent aggregation or undesired particle growth. Unfortunately for many applications, the organic groups do not have useful electrical or optical properties. There has thus been much effort to replace the organic groups with inorganic groups that either add to the capabilities of the crystals or can be converted into an electrically or optically active material. This approach has been used to make nanocrystal assemblies with greatly improved electrical properties⁴ and to convert nanocrystals capped with inorganic complexes into a more useful photovoltaic material⁵ (a material that converts light into electricity).

Llordés *et al.* have used this strategy to create their nanoparticle-in-glass materials. The authors first stripped indium tin oxide (ITO) nanocrystals of their organic caps and replaced them with niobium-containing polyatomic ions known as polyoxometalate (POM) clusters. These clusters attach covalently to

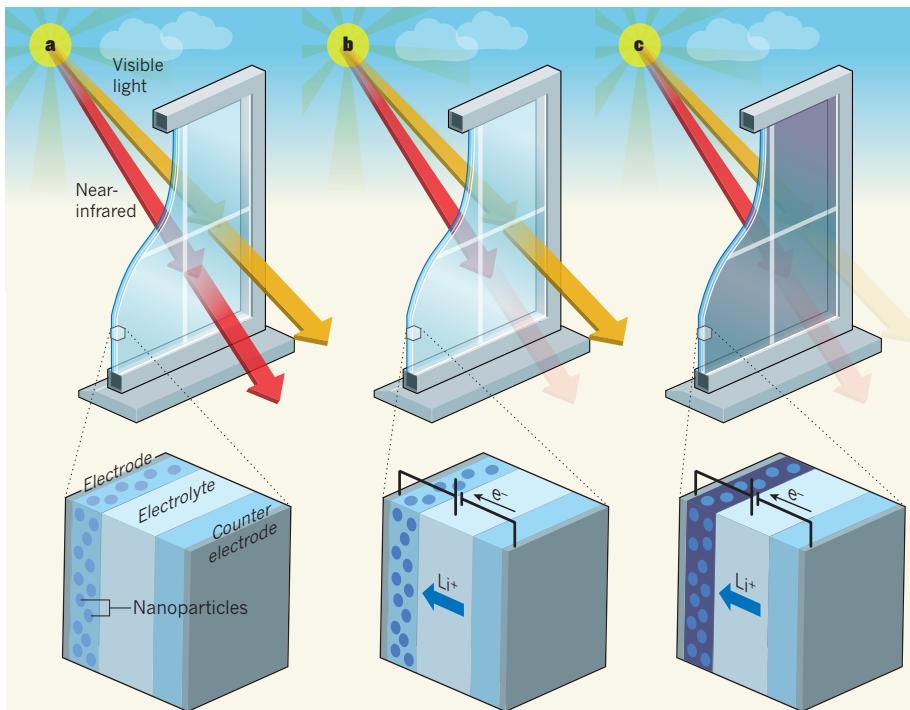


Figure 1 | Electrochromic window design. Llordés *et al.*³ propose that their nanoparticle-in-glass composite material could be used to make windows that controllably and selectively absorb visible light and near-infrared light (heat). **a**, In the design, the window is an electrochemical cell in which two conducting glass panes are separated by a solid electrolyte material. The authors' material is deposited on one pane, forming an electrode; a counter electrode is deposited on the other pane. In the absence of an electrical load, the window is transparent to visible and near-infrared light. **b**, When an intermediate voltage is applied, charge carriers (lithium ions, Li^+ , and electrons, e^-) move through the circuit. The nanoparticles in the composite become chemically reduced, whereupon they block most incoming near-infrared light. **c**, At lower voltages, the glass matrix of the composite also becomes reduced and blocks most incoming visible light.

the ITO surface to create a shell around the nanocrystal. The researchers then condensed the modified nanocrystals into a film, simply by evaporating the solvent from a dispersion of the crystals. Finally, they converted the POM between the densely packed ITO nanocrystals into a niobium oxide (NbO_x) glass matrix by heating the film to 400 °C. Compared with previously reported synthetic routes for making nanoparticle-in-glass materials, in which inorganic crystals are grown within a glass⁶, Llordés and co-workers' method provides rigorous control over the nanocrystals' size distribution and volume fraction. And, by adding more POM to the dispersion of POM-stabilized ITO nanocrystals, the authors could

increase the volume fraction of the NbO_x glass matrix.

One of the key features of the ITO nanocrystal– NbO_x glass material is that the glass is covalently bonded to the nanocrystals. This restricts the molecular orientations available to the octahedral NbO_6 units found in the glass, and leads to remarkable structural ordering that differs from that of pure NbO_x . It turns out that this ordering improves the electrochromic properties of the glass matrix: NbO_x in the composite is five times darker than the bulk material when a similar voltage is applied.

ITO nanocrystals are also electrochromic, but in a different wavelength region from NbO_x ; they undergo reversible electrochemical

redox reactions and absorb near-infrared light in the reduced state, but are transparent to this part of the spectrum when oxidized⁷. The combination of ITO nanocrystals with a NbO_x glass matrix therefore yields a material in which both visible and near-infrared light absorption can be electrochemically modulated. This material could thus be used in smart windows, to control the amount of both heat (near-infrared) and light passing through them (Fig. 1). What's more, the optical transparency can be tuned independently of the near-infrared transparency.

Llordés and colleagues' approach for making composite materials of inorganic nanocrystals in glass opens the way to a range of new material properties and applications, not just in electrochromics. The challenge for each application is to identify the best combinations of nanocrystal composition and modifiable inorganic capping groups.

More specifically, several issues must still be addressed before the material can be used in windows. The authors used lithium metal as a counter electrode to test the performance of their material, but this is not acceptable for commercial applications because of safety concerns. A suitable counter electrode must be identified. Additionally, the researchers performed their photoelectrochemical tests using a liquid electrolyte as a charge-carrying material, whereas a solid electrolyte is probably more appropriate for buildings applications. The materials needed to build an electrochromic window will be more expensive than conventional window materials, so the extra expense will need to be balanced by the energy and cost savings that can be achieved through their use. Ideally, no power input will be needed to maintain transparency or opacity, but this ability remains to be explored.

Nevertheless, Llordés and co-workers' results are promising. With appropriate counter electrodes and a solid-state electrolyte, and if long-term stability of the composite can be demonstrated, windows that have multispectral band transparency may be just around the corner, potentially enabling buildings that offer unprecedented energy efficiency and comfort. ■

Brian A. Korgel is in the Department of Chemical Engineering, Center for Nano- and Molecular Science and Technology, Texas Materials Institute, The University of Texas at Austin, Austin, Texas 78712, USA.
e-mail: korgel@mail.che.utexas.edu

1. Richter, B. et al. *Rev. Mod. Phys.* **80**, S1–S109 (2008).
2. Li, S.-Y., Niklasson, G. A. & Granqvist, C. G. *J. Appl. Phys.* **108**, 063525 (2010).
3. Llordés, A., García, G., Gazquez, J. & Milliron, D. J. *Nature* **500**, 323–326 (2013).
4. Panthani, M. G. & Korgel, B. A. *Annu. Rev. Chem. Biomol. Eng.* **3**, 287–311 (2012).
5. Jiang, C., Lee, J.-S. & Talapin, D. V. *J. Am. Chem. Soc.* **134**, 5010–5013 (2012).
6. Sakamoto, A., Yamamoto, S. *Int. J. Appl. Glass Sci.* **1**, 237–247 (2010).
7. García, G. et al. *Nano Lett.* **11**, 4415–4420 (2011).

STRUCTURAL BIOLOGY

RNA exerts self-control

A crystal structure of two bound RNA molecules not only provides insight into how regulatory riboswitch sequences affect messenger RNA expression, but also expands our understanding of RNA structure and architecture. SEE LETTER P.363

BHASKAR CHETNANI &
ALFONSO MONDRAGÓN

Over the past three decades, our knowledge of the role of RNA in cellular processes has expanded enormously¹. For example, the structures of various regulatory RNA sequences called riboswitches have revealed how they affect the transcription or translation of their downstream messenger RNA sequences through the recognition and binding of specific ligands². On page 363 of this issue, Zhang and Ferré-D'Amare³ provide another structural insight — this time, into a bacterial riboswitch, called a T-box, in complex with a transfer RNA molecule. Their data elucidate how one RNA molecule recognizes another RNA of similar size and regulates its own transcription through a simple switching mechanism*.

To support the process of protein synthesis, cells must regulate the pool of tRNAs that become charged with (covalently bound to) specific amino-acid residues and deliver them to the growing protein chain. Enzymes known as aminoacyl-tRNA synthetases carry out tRNA charging. In Gram-positive bacteria, a T-box riboswitch located upstream of the coding region of the mRNA of each

aminoacyl-tRNA synthetase negatively regulates synthesis of the mRNA⁴.

The T-box RNA consists of at least two independently folded domains: a sensory 'aptamer' domain that forms a long stem called stem I and that binds to specific tRNAs; and a second domain, which can switch between two alternative conformations depending on whether the bound tRNA is charged or uncharged⁵. Whereas binding of a charged tRNA leads to termination of transcription of the aminoacyl-tRNA coding sequence (Fig. 1a), an uncharged tRNA stabilizes an antiterminator conformation of the T-box, leading to transcription of the mRNA and subsequent protein synthesis (Fig. 1b).

Stem I is a mostly double-helical RNA domain roughly 100 nucleotides long and is studded with several phylogenetically conserved structural motifs along its length⁴. Previous work has implicated most of these motifs in tRNA recognition and binding⁴, yet atomic-level details of this process have remained largely unknown. Zhang and Ferré-D'Amare provide the first glimpse into this mechanism by describing the crystal structure of a complex between stem I of the T-box of tRNA synthetase for the amino acid glycine and an uncharged glycyl-tRNA. The authors' co-crystal structure also explains the precise role of the stem I motifs, delivers intriguing

*This article and the paper under discussion³ were published online on 28 July 2013.

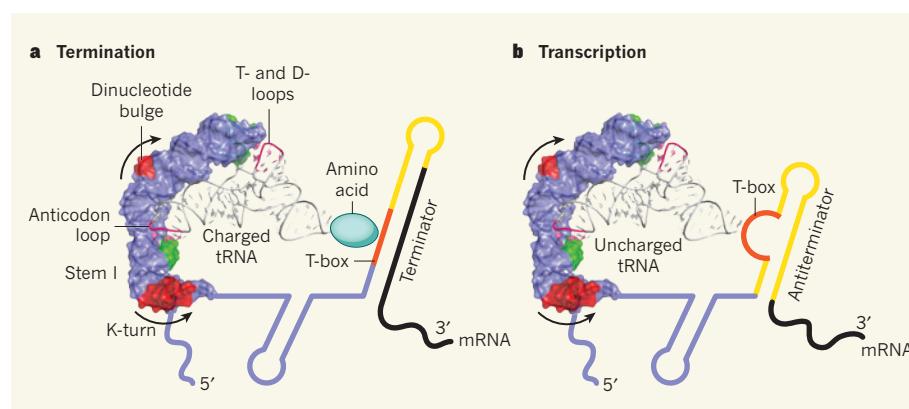


Figure 1 | RNA meets RNA. Zhang and Ferré-D'Amare³ present the structure of the stem I domain of a T-box riboswitch in complex with an uncharged tRNA. Stem I bends at the dinucleotide bulge and the K-turn to recognize tRNA by binding to its anticodon loop and the T- and D-loops. **a**, The amino acid present in a charged tRNA is thought to prevent interaction between the T-box sequence motif downstream of stem I and the acceptor end of the tRNA, favouring formation of a terminator loop that stops transcription of the mRNA downstream of the T-box. **b**, By contrast, the free acceptor end of an uncharged tRNA interacts directly with the T-box sequence and leads to the formation of an antitermination loop, allowing transcription of the mRNA and its subsequent translation into a protein.

information about their synergy and offers a structural rationale for their sequence conservation.

Specific recognition of a tRNA by stem I seems to involve two main tRNA regions: the anticodon loop and the T- and D-loops. Specifically, stem I bends to follow closely the shape of the tRNA and contacts the anticodon loop and the T- and D-loops with its proximal and distal ends, respectively.

The interaction between the tRNA T- and D-loops and the T-box stem I was anticipated, on the basis of recent bioinformatics, biochemical and structural data^{6,7}. For example, a previous crystal structure of a distal segment of the same stem I showed⁷ two interleaved loops forming a similar arrangement to the one observed in two other large and unrelated RNAs that also recognize tRNAs — the ribozyme RNase P and an RNA component of the ribosomes^{8,9}. In those large RNAs, the interleaved loops participate in tRNA recognition and binding by stacking their planar bases over evolutionarily conserved unstacked bases in the tRNA.

Zhang and Ferré-D'Amare's structure of stem I and tRNA corroborates these predictions and demonstrates conclusively that the interleaved loops recognize the conserved tRNA 'elbow' in a similar fashion to that in the previously observed cases^{8,9}. The structure also reveals a surprise — the interactions between the three-nucleotide anticodon sequence of glycyl-tRNA and stem I are almost identical to those made between a tRNA anticodon and mRNAs on the ribosome^{9,10}.

The present paper further shows an elegant 'mutually induced fit' mechanism for T-box riboswitches, by which both binding partners change conformation to attain shape complementarity. Finally, whereas the inherent flexibility of a tRNA is well established¹¹, the co-crystal structure demonstrates how stem I bends sharply around two hinge regions to embrace the tRNA. The hinge regions comprise a conserved dinucleotide bulge and K-turn motif, respectively (Fig. 1), and the structure explains why mutations in these hinges lead to impaired tRNA binding and regulation^{3,12}.

Zhang and Ferré-D'Amare's work is thus a big step towards understanding the mechanism of action of T-box riboswitches at the atomic level. It explains how a relatively small RNA molecule can recognize a specific tRNA by maximizing interactions. Details of the other important parts of the switching mechanism — namely, recognition of the charged state of the tRNA and the conformational changes that lead to regulation — remain unknown, but this study brings us closer to understanding this fascinating riboswitch in even greater detail.

An added bonus is that the authors' structure provides truly interesting information on RNA–RNA recognition and on RNA structure

and architecture. Despite the paucity of structures of large RNA molecules, some commonalities are starting to emerge. In all known cases, recognition of tRNA involves at least two distinct and distant areas. RNA flexibility plays a notable part in recognition and serves to maximize interactions and achieve shape complementarity. In addition, the recurrent use of a few structural motifs seems to have a large role in RNA architecture, as illustrated by the presence of several previously known structural motifs along stem I.

The structure of the T-box stem I–tRNA complex is one more example of the conformational versatility of RNA that allows it to perform various functions. As we learn more about RNA structure and function, it frequently emerges that RNA molecules can execute many of the functions that are normally ascribed to proteins. One can, therefore, imagine that there was a time when many of the functions that are now performed by proteins or by proteins in complex with nucleic acids were carried out solely by versatile RNA molecules. We look forward to more examples

of RNA molecules having unexpected roles in biology. ■

Bhaskar Chetnani and Alfonso Mondragón
are in the Department of Molecular
Biosciences, Northwestern University,
Evanston, Illinois 60208, USA.
e-mail: a-mondragon@northwestern.edu

- Atkins, J. F., Gesteland, R. F. & Cech, T. *RNA Worlds* (Cold Spring Harbor Laboratory Press, 2011).
- Serganov, A. & Nudler, E. *Cell* **152**, 17–24 (2013).
- Zhang, J. & Ferré-D'Amare, A. R. *Nature* **500**, 363–366 (2013).
- Green, N. J., Grundy, F. J. & Henkin, T. M. *FEBS Lett.* **584**, 318–324 (2010).
- Yousef, M. R., Grundy, F. J. & Henkin, T. M. *J. Mol. Biol.* **349**, 273–287 (2005).
- Lehmann, J., Jossinet, F. & Gautheret, D. *Nucleic Acids Res.* **41**, 5494–5502 (2013).
- Grigg, J. C. et al. *Proc. Natl Acad. Sci. USA* **110**, 7240–7245 (2013).
- Reiter, N. J. et al. *Nature* **468**, 784–789 (2010).
- Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. F. *Cell* **126**, 1065–1077 (2006).
- Selmer, M. et al. *Science* **313**, 1935–1942 (2006).
- Dethoff, E. A., Petzold, K., Chugh, J., Casiano-Negrón, A. & Al-Hashimi, H. M. *Nature* **491**, 724–728 (2012).
- Winkler, W. C., Grundy, F. J., Murphy, B. A. & Henkin, T. M. *RNA* **7**, 1165–1172 (2001).

GENE THERAPY

Primed for take-off

Gene therapy is finally getting a bumper crop of data that show clinical efficacy after fine-tuning of key parameters that control safety and potency. Supporting evidence comes from treatment of two life-threatening human diseases.

PHILIPPE LEBOUCH

The development of the field of gene therapy shares many similarities with the history of aviation. Each is based on deceptively simple principles: the introduction of a therapeutic gene into cells and the flow of air over an aircraft's wing. However, the devil is in the details, and the details are in the numbers. Each field was marred by shortcomings and adverse events early on. But in spite of naysayers lacking vision, both fields continued their quest, and there is now firm hope that gene therapy will soon do for medicine what aeroplanes did for transportation. Good reason for this optimism comes from the results of two studies published in *Science* by Biffi *et al.*¹ and Aiuti *et al.*².

Both studies make use of lentiviral vector transduction, which continues to pave the way for gene therapy³. The technique involves collecting a type of stem cell from the body, adding therapeutic genes using a delivery system (or vector) that has been safely engineered from a few components of HIV, and putting the modified cells back into the patient. The first international trials using lentiviral vectors for the gene therapy of inherited disorders were

conducted in France for the diseases adrenoleukodystrophy⁴ and β-thalassaemia⁵, and provided evidence of clinical efficacy in several patients for at least six years. The *Science* papers report the results of two more trials — one for metachromatic leukodystrophy (MLD)¹ and another for Wiskott–Aldrich syndrome² — with further quantitative improvements in efficacy.

MLD is an inherited disorder caused by deficiency of the enzyme arylsulphatase A (ARSA). Patients with MLD accumulate a toxic metabolite in the nervous system, which results in massive cognitive and motor impairments. Children with the severe form, called MLD-LI, die within a few years of symptom onset, and current treatments are ineffective.

Biffi *et al.* collected blood-making cells called haematopoietic stem cells (HSCs) from the bone marrow of three children with MLD-LI who were not yet showing symptoms, and exposed the cells to an ARSA-expressing lentiviral vector. These transduced cells were then re-injected into the children, who meanwhile had received a brief course of chemotherapy to eliminate uncorrected natural HSCs. This process is known as autologous

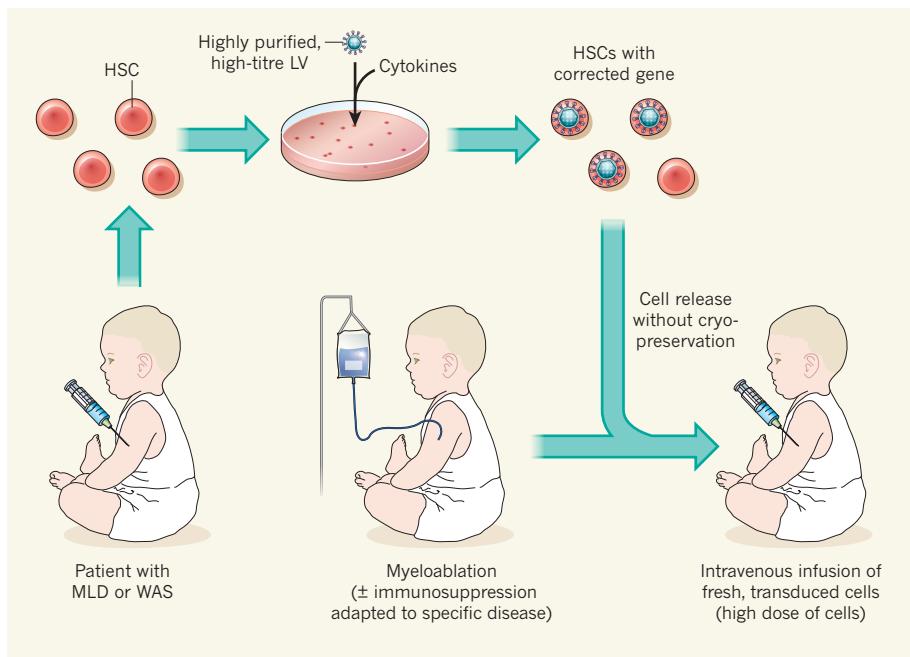


Figure 1 | Life-saving vectors. Researchers^{1,2} harvested haematopoietic stem cells (HSCs) from the bone marrow of children with mutations in either the gene that causes metachromatic leukodystrophy (MLD) or the gene that results in Wiskott–Aldrich syndrome (WAS). In the lab, they then exposed the derived HSCs to highly purified and concentrated lentiviral vectors (LV) carrying the normal genes in the presence of cytokine molecules to stimulate vector integration into HSC genomic DNA. Before re-injecting the vector-carrying cells into the patients' blood, the authors subjected the children to myeloablative chemotherapy (and immunosuppression in the case of WAS) to destroy their remaining defective HSCs and allow successful incorporation of the genetically corrected HSCs. Indicated are some of the steps that resulted in improvement in gene-transfer efficiency and clinical benefit over previous gene-therapy trials.

transplantation with myeloablative conditioning (Fig. 1). The authors hypothesized that blood cells now expressing the missing enzyme at above-physiological levels (specific enhancer and promoter sequences were used to drive efficient expression of ARSA in the lentiviral vector) would bring enough of the missing enzyme to the nervous system, directly or indirectly, to have therapeutic benefit.

The results were beyond expectations. Almost two years after the therapy, more than 60% of blood cells contained the vector, with several vector copies per cell, resulting in ARSA levels ten times greater than in healthy individuals. These above-physiological levels are desirable for optimal potency and are considered safe. The authors also detected ARSA activity in the cerebrospinal fluid, and there was no clinical progression of the disease in any of the three patients 7–21 months after the predicted age of symptom onset.

Wiskott–Aldrich syndrome is also an inherited disorder, in which a gene defect causes immunodeficiency. Focusing on three children with the disorder, Aiuti and colleagues used a similar approach to that of Biffi *et al.*, but chose a different conditioning regimen. Because corrected HSCs from patients with this syndrome are known to have a substantial selective advantage *in vivo*, the authors used

less myeloablation but more immunosuppression to stimulate engraftment of the genetically modified immunocompetent cells.

Aiuti *et al.* also observed high levels (25–50%) of blood cells expressing the lentiviral vector in all three patients. Although quantitative interpretation of this result is difficult because of the type of conditioning regimen used and the existence of substantial *in vivo* selection, this clinical response is impressive and better than that achieved in previous gene-therapy trials for Wiskott–Aldrich syndrome⁶.

What made these striking improvements possible? First, lentiviral vectors continue to rule unsurpassed because of their many useful properties. These include effective transduction of quiescent HSCs; compatibility with therapeutic gene constructs of large size and complexity; effective concentration and purification of vector particles to increase viral titres and potency; and amenability to cryopreservation³ (Fig. 1). But as the authors suggest^{1,2}, the answer may just be a numbers game.

Vector copy numbers in patients' blood cells after engraftment in the MLD trial were around five times higher than those observed in trials for adrenoleukodystrophy⁴ and β-thalassaemia⁵ (around 0.2 copies of the vector after 6 years; my unpublished data). Moreover, the latest studies may have

benefited from a higher degree of vector purity following its high-titre production. This would have minimized contamination from defective interfering vector particles. Another potentially important modification was the immediate reinfusion of non-cryopreserved cells, thus avoiding compromising HSCs before their engraftment. The very young age of the patients (7–16 months for MLD compared with 18 years for β-thalassaemia⁵) and the large dose of the infused cells may also have contributed positively.

What about safety? The main concern with HSC gene therapy is the potential activation of cancer genes following random integration of vectors in the host genome, because efficient tethering of lentiviral vectors to genomic areas of lesser cancer risk is not currently feasible in these cells. Biffi and colleagues and Aiuti *et al.* take comfort from the fact that their vectors showed a heterogeneous pattern of integration into chromosomal sites, as observed after pan-genomic monitoring of blood cells. This is a presumptive indicator that no particular genomic areas carrying the vector may result in pre-malignant outgrowth of an HSC or another blood-cell clone.

Yet the meaning of such analyses remains unclear⁷ and can either be falsely reassuring or alarming. For instance, the partially dominant cell clone bearing an integration site within the *HMGA2* gene in a β-thalassaemia patient raised anxiety among observers about the possibility of cancer development⁵. But six years after gene therapy, this clone is no longer dominant, although vector copy numbers in the patient's cells and expression of the corrected gene product, β-globin, remain sufficiently high for the patient to be completely independent of blood transfusions (my unpublished data). It is intriguing that *HMGA2* has now been identified as a vector integration site in various studies, including the present trial for Wiskott–Aldrich syndrome².

The time is ripe for pharmaceutical companies to grab the baton and bring gene therapies to patients in need. The naysayers now whisper that gene therapy is too expensive to treat patients, no matter where they are. But compared with the enormous lifelong cost of enzyme replacement or even taking small-molecule drugs, the thoughtful pricing of many gene-therapy approaches that require a one-off treatment may meet the respective requirements of patients, the pharmaceutical industry and institutional funders, even in developing countries. ■

Philippe Leboulch is at the Institute of Emerging Diseases and Innovative Therapies, CEA, INSERM, University of Paris 11, 92265 Fontenay-aux-Roses, France. He is also at the Brigham & Women's Hospital and Harvard Medical School, Boston, Massachusetts, USA.
e-mail: pleboulch@rics.bwh.harvard.edu

1. Biffo, A. et al. *Science* **339**, 1233158 (2013).
2. Aiuti, A. et al. *Science* **339**, 1233151 (2013).
3. Naldini, L. *Nature Rev. Genet.* **12**, 301–315 (2011).
4. Cartier, N. et al. *Science* **326**, 818–823 (2009).
5. Cavazzana-Calvo, M. et al. *Nature* **467**, 318–322 (2010).
6. Bozta, K. et al. *N. Engl. J. Med.* **363**, 1918–1927 (2010).
7. Tubsuwan, A. et al. *Stem Cells* **31**, 1436 (2013). doi:10.1002/stem.1436 (2013).

QUANTUM COMMUNICATION

Reliable teleportation

Two complementary experiments have demonstrated deterministic quantum teleportation of quantum bits. The results could find applications in quantum communications and computing. SEE LETTERS P.315 & P.319

TIMOTHY C. RALPH

The ability to teleport quantum states from one place to another by means of classical communication systems is both counter-intuitive and potentially useful. However, experimental demonstrations of such quantum teleportation have tended to be hit-and-miss, with few successful attempts. Two studies in this issue now report deterministic quantum teleportation: one involves the teleportation of the quantum state of a single photon, by Takeda *et al.*¹ (page 315); the other is teleportation of the quantum state of a superconducting circuit, by Steffen *et al.*² (page 319).

Quantum teleportation³ relies on the ability to produce and measure entangled states. Entanglement refers to strong correlations between distinct quantum systems that defy classical explanation. Generically, teleportation protocols proceed by three steps (Fig. 1): a pair of quantum systems in an entangled state is produced and distributed, one to a sender (Alice) and the other to a receiver (Bob); Alice makes a joint measurement of her member of

the entangled pair and the unknown state she wishes to teleport, and sends the measurement result to Bob; Bob uses the measurement result he receives from Alice to manipulate his quantum system in a predetermined way. After this manipulation, Bob's quantum system ends up being in the unknown state, that teleported from Alice to Bob, with the only direct communication being a classical message — Alice's measurement result.

Most previous quantum teleportation experiments have been non-deterministic because some steps of the protocol, in particular producing the initial entanglement and carrying out the joint measurement, have been difficult to perform efficiently. For example, in single-photon quantum optics, which most earlier teleportation studies have used, the following restrictions apply: first, entangled pairs are produced by a spontaneous process that typically succeeds less than once in 100 attempts; second, with linear optics, the required joint measurement can be performed with a probability of success of only 50%. In their studies, Takeda *et al.* and Steffen *et al.*

take very different approaches to overcoming this problem of limited efficiency.

Takeda and colleagues' experiment teleports a single-photon quantum bit (qubit) with much higher efficiency than all earlier optical studies. A qubit is a quantum system that has two distinct outcomes — in this case, a single photon in a superposition of two different arrival times at a detector. The trick the authors use is to perform a more general form of teleportation that works for any optical-field state, rather than only for single-photon qubit states. This sounds, and is, hard to achieve, but it turns out to make a big difference because deterministic sources of optical-field entanglement and efficient joint field measurements are both available. This 'continuous-variable' teleportation protocol⁴ has previously been demonstrated for multi-photon states^{5,6}, but this is the first time that a single-photon qubit has been teleported by this method. The result is the deterministic teleportation of a single-photon qubit with a quality that exceeds the limits set by any classical protocol — that is, one that does not use entanglement but in which Alice instead tries to directly measure (necessarily imperfectly) the state of the qubit and sends this information to Bob.

Steffen and colleagues tackled the efficiency problem in a different way. They implemented teleportation of a solid-state qubit — a superconducting circuit. The qubit circuit is on the 100-micrometre scale and is held at temperatures of around 20 millikelvin. Qubits are formed from small currents running in the circuit⁷, and at these sizes and temperatures, the currents behave quantum mechanically. As in Takeda and colleagues' study, photons are still used to move information around, but these are now at microwave (rather than optical) frequencies⁸. Alice and Bob are separated by about 5 millimetres. Strong interactions between the superconducting qubits and the microwave photons allow entanglement to be produced deterministically between the superconducting qubits held by Alice and Bob. Such interactions can also be used to make a joint measurement deterministically between Alice's qubit and the unknown qubit, as required by the teleportation protocol. Again, the result is deterministic teleportation with a quality that exceeds the classical limit. Deterministic teleportation has been achieved previously in ion-trap experiments, but in that situation the distances between the qubits were 1,000 times smaller^{9,10}.

Both experiments are 'hero' experiments, in the sense that they only just exceed classical limits when they are running in the fully deterministic mode. In Takeda and colleagues' study, this is mainly due to the limited strength of the field entanglement, whereas for Steffen and co-workers' study the main problem is imperfect differentiation of the outcomes of the joint-measurement results. However, both groups are also able to run their experiments in

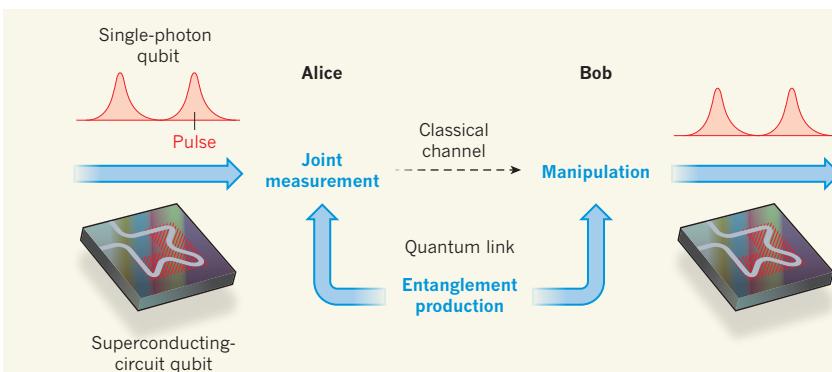


Figure 1 | A quantum teleportation protocol. A qubit, here either a single photon¹ in a superposition of two pulses or a superconducting circuit², is prepared in an unknown state. A pair of quantum systems in an entangled state has previously been prepared and distributed between a sender (Alice) and a receiver (Bob). Alice makes a joint measurement between the unknown state, which she wishes to teleport, and her member of the entangled pair. The measurement result is sent to Bob via a classical channel. He uses it to manipulate his member of the entangled pair, in the process recreating the unknown state at his station. The quantum link that allows the entangled pair to be distributed to Alice and Bob is optical in the case of the photon and uses microwave photons in the superconducting case.

a non-deterministic mode. In Takeda and colleagues' experiment, the photon qubit can be successfully teleported about 40% of the time (compared with a value of much less than 1% for all previous optical studies), and reproduced at Bob with about 88% quality (similar to the best achieved in previous experiments). Steffen *et al.* can arrange for their qubit to be teleported 25% of the time with a quality of about 82%. For both teams, the limitations in their experiments are clearly understood and shown not to be fundamental impediments to future improvements.

The advances made in these experiments should flow on to allow improved quantum-information protocols. Takeda and co-workers' optical 'flying qubits' have potential applications in quantum communications. But there is a caveat. Improving quantum communications using teleportation requires the purification of entangled states sent through a noisy channel. Teleportation could therefore be used to transfer quantum states between distant locations with better quality than sending them directly. Purification can be achieved by distillation techniques. However, distillation techniques for field entanglement are not as advanced as those for qubit entanglement. Nevertheless, promising advances in distilling field entanglement have been made¹¹. By contrast, the solid-state 'standing qubits' in Steffen and colleagues' experiment are more likely to find applications in quantum computing¹². Notably, they demonstrate the increasing sophistication and quality of the manipulations possible with superconducting qubits coupled to microwave transmission lines, and raise the profile of such qubits as potential building blocks for large-scale quantum computation.

More progress is needed before deterministic quantum teleportation under practical conditions, and with quality approaching 100%, becomes a reality. But these experiments represent significant steps along that path. ■

Timothy C. Ralph is in the Centre for Quantum Computation and Communication Technology, School of Mathematics and Physics, University of Queensland, Brisbane, Queensland 4072, Australia.
e-mail: ralph@physics.uq.edu.au

1. Takeda, S., Mizuta, T., Fuwa, M., van Loock, P. & Furusawa, A. *Nature* **500**, 315–318 (2013).
2. Steffen, L. *et al.* *Nature* **500**, 319–322 (2013).
3. Bennett, C. H. *et al.* *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
4. Braunstein, S. L. & Kimble, H. J. *Phys. Rev. Lett.* **80**, 869–872 (1998).
5. Furusawa, A. *et al.* *Science* **282**, 706–709 (1998).
6. Bowen, W. P. *et al.* *Phys. Rev. A* **67**, 032302 (2003).
7. Clarke, J. & Wilhelm, F. K. *Nature* **453**, 1031–1042 (2008).
8. Wallraff, A. *et al.* *Nature* **431**, 162–167 (2004).
9. Riebe, M. *et al.* *Nature* **429**, 734–737 (2004).
10. Barrett, M. D. *et al.* *Nature* **429**, 737–739 (2004).
11. Xiang, G. Y. *et al.* *Nature Photon.* **4**, 316–319 (2010).
12. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* (Cambridge Univ. Press, 2000).

SYSTEMS BIOLOGY

Metabolite turns master regulator

The phenomenon of catabolite repression enables microorganisms to use their favourite carbon source first. New work reveals α -ketoacids as key effectors of this process, with their levels regulating gene expression. SEE ARTICLE P.301

JOSHUA D. RABINOWITZ &
THOMAS J. SILHAYV

Nutrients in the environment are a primary determinant of microbial physiology. When preferred nutrients are abundant, microbes grow fast. When they are scarce, growth slows down. This change in growth rate is accompanied by a change in cellular composition, with fast-growing cells being loaded with ribosomes (which are needed for rapid protein production), and slower-growing cells containing more metabolic enzymes for nutrient assimilation (catabolism)^{1,2}. In this issue, You *et al.* (page 301)³ identify a striking linear relationship between the total protein composition (the proteome) of a cell and its growth rate, which extends beyond ribosomes to metabolic enzymes*. They further demonstrate how such a relationship can arise, in part, from a new regulatory connection, in which a particular class of carbon catabolite called α -ketoacids, which form

*This article and the paper under discussion³ were published online on 7 August 2013.

the carbon skeletons of amino acids, serves as a master transcriptional regulator by inhibiting the production of cyclic AMP — the primary inducer of carbon-catabolic genes.

Perhaps the most intensively studied example of gene regulation involves the enzymes of the bacterium *Escherichia coli* that mediate lactose catabolism. These enzymes are expressed only when lactose is present and glucose (the preferred carbon source) is not⁴. *Escherichia coli* detects the presence of lactose through binding of this sugar to, and inactivation of, the lac repressor protein⁵, and it senses the absence of glucose from elevated levels of cAMP⁶, which binds to and activates the transcription factor Crp (refs 7, 8). Identification of this classic regulatory loop involved seminal contributions from three Nobel laureates — François Jacob and Jacques Monod, who won the 1965 physiology prize for their pioneering studies of gene regulation⁵, and Earl Sutherland, winner of the 1971 prize for discoveries related to cAMP, who later identified cAMP in *E. coli* and showed that its levels rapidly fall in response to glucose⁶.

What controls the activity of adenylate

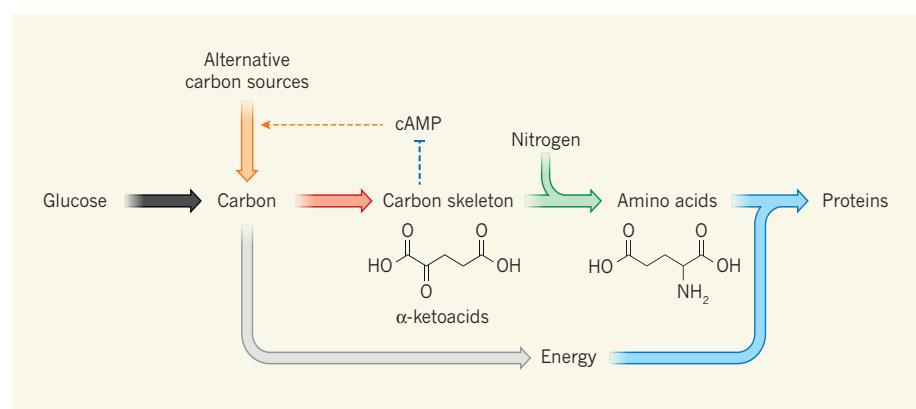


Figure 1 | Regulation of cAMP levels by carbon and nitrogen availability. The main biosynthetic task of the bacterium *Escherichia coli* is protein production. This requires energy generation (grey arrow), carbon-skeleton synthesis, nitrogen incorporation to make amino acids, and protein synthesis. The anabolic fluxes (red, green and blue arrows) increase linearly with growth rate, as does the required anabolic proteome fraction except for processes that are directly slowed by nutrient limitation (such as amino-acid synthesis during nitrogen limitation). Carbon catabolism provides energy and building blocks for anabolism, and cAMP allocates the proper fraction of the proteome to carbon-catabolic enzymes. You *et al.*³ find that this is achieved by a new regulatory loop, wherein α -ketoacids inhibit cAMP production: when favoured carbon sources such as glucose are present or nitrogen is limiting, carbon influx exceeds anabolic capability and α -ketoacid accumulation inhibits cAMP. Conversely, when favoured carbon sources are depleted, α -ketoacid levels fall, and cAMP increases to stimulate production of the required carbon-catabolic machinery (orange arrows).

cyclase (the enzyme that makes cAMP) and so cAMP levels? The phosphorylated form of the enzyme EI^AGlc can activate adenylate cyclase. EI^AGlc is a component of the bacterial phosphotransferase system, which takes the phosphate group from phosphoenolpyruvate — the last intermediate in the biochemical process of glycolysis — and passes it through a series of enzymes, eventually leading to glucose import, phosphorylation and metabolism. When glucose is absent, the phosphorelay activity of EI^AGlc ceases, and the phosphorylated enzyme induces cAMP production⁹.

Although elegant, this is not the full story. Carbon sources that are not imported by way of the phosphotransferase system also tend to counteract cAMP production. In 1961, Boris Magasanik hypothesized that the general ability of carbon sources to repress the expression of catabolic enzymes reflects the fact that all carbon sources converge to produce a key signalling metabolite — a process he termed catabolite repression¹⁰. The relevant catabolites, however, were never identified.

You *et al.* now show that the key catabolites are α-ketoacids, which inhibit adenylate cyclase independently of the phosphotransferase system. Although several α-ketoacids can inhibit adenylate cyclase, α-ketoglutarate is the most abundant¹¹ and therefore likely to be physiologically dominant. As both an intermediate of the energy-producing tricarboxylic acid cycle and the carbonaceous substrate for nitrogen assimilation, α-ketoglutarate reflects the balance of carbon to nitrogen in available nutrients. Inhibition of adenylate cyclase by α-ketoglutarate explains the long-standing observations^{12–14} that low nitrogen availability blocks expression of what we now know are cAMP-induced genes. Fascinatingly, the role of α-ketoglutarate as both a metabolite and a master regulator is evolutionarily conserved; in humans it serves as a cofactor to enzymes that covalently modify transcription factors, histone proteins and DNA.

How and why does this regulatory connection lead to linear relationships between gene expression and cellular growth rate in microbes? When carbon limitation slows cellular growth, there is an increasing need for carbon-catabolic enzymes and a decreasing need for anabolic enzymes, which promote biosynthesis and use carbon as a building block (Fig. 1). Quantitatively, in the absence of futile cycling (in which two metabolic pathways operating in opposite directions cancel out each other's effects, wastefully using up energy), the cell's anabolic metabolism is directly proportional to growth rate. So, as growth slows, the required amount of anabolic enzymes, assuming their activities are constant, decreases linearly. Accordingly, the resulting 'empty space' in the proteome — which will be filled with enzymes required for coping with the carbon limitation, such as cAMP-regulated gene products — increases

linearly with decreasing cellular growth.

When growth slows because of nitrogen limitation, there is less need for carbon-catabolic enzymes and more demand for those involved in nitrogen assimilation. This time, the decrease in the requirement for carbon-catabolic flux is linear with decreasing growth rate. Thus, the situation is flipped, but the optimal responses still remain linear.

Inhibition of cAMP production by α-ketoglutarate naturally produces the desired responses. Whenever carbon-catabolic machinery is in excess relative to anabolic machinery, α-ketoglutarate accumulates, cAMP levels fall and carbon-catabolic enzymes are repressed. Conversely, when anabolic machinery is in excess, α-ketoglutarate is depleted, cAMP levels rise and carbon-catabolic enzymes increase. The steady-state concentration of cAMP is therefore the factor that ensures that the proper amount of the proteome is devoted to carbon-catabolic enzymes. Consequently, the physiological function of cAMP signalling goes beyond simply enabling hierarchical utilization of carbon sources, which is of unclear significance for fitness and can also be achieved through other mechanisms (including inducer exclusion¹⁵). Instead, cAMP controls the fraction of the proteome devoted to carbon catabolism.

In a broader historical context, during the half-century between the coining of the term catabolite repression and the present work, physiology has largely taken a back seat to molecular genetics and, more recently, to genomics. Therefore, You and colleagues' use of quantitative physiology to elucidate the molecular mechanism of catabolite repression is particularly noteworthy. One hopes that it is

the beginning of a rebalancing, in which physiology-driven systems biology emerges as a full equal to research driven by molecular biology. As the present paper shows, the strength of the physiology-driven approach is not in finding the full scope of molecular events occurring in biological systems, but in identifying the most functionally important ones. ■

Joshua D. Rabinowitz is in the Department of Chemistry and at the Lewis-Sigler Institute for Integrative Genomics, Princeton University, and **Thomas J. Silhavy** is in the Department of Molecular Biology, Princeton University, Princeton, New Jersey 08544, USA.
e-mail: joshr@genomics.princeton.edu

- Maaløe, O. in *Biological Regulation and Development* (ed. Goldberger, R. F.) 487–542 (Plenum, 1979).
- Klumpp, S., Zhang, Z. & Hwa, T. *Cell* **139**, 1366–1375 (2009).
- You, C. *et al.* *Nature* **500**, 301–306 (2013).
- Monod, J. *Recherches sur la croissance des cultures bactériennes*. PhD thesis (Univ. Paris, 1942).
- Jacob, F. & Monod, J. *J. Mol. Biol.* **3**, 318–356 (1961).
- Makman, R. S. & Sutherland, E. W. *J. Biol. Chem.* **240**, 1309–1314 (1965).
- Emmer, M., deCrombrugge, B., Pastan, I. & Perlman, R. *Proc. Natl Acad. Sci. USA* **66**, 480–487 (1970).
- Zubay, G., Schwartz, D. & Beckwith, J. *Proc. Natl. Acad. Sci. USA* **66**, 104–110 (1970).
- Deutscher, J., Francke, C. & Postma, P. W. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).
- Magasanik, B. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 249–256 (1961).
- Bennett, B. D. *et al.* *Nature Chem. Biol.* **5**, 593–599 (2009).
- Mandelstam, J. *Nature* **179**, 1179–1181 (1957).
- Daniel, J. & Danchin, A. *Biochimie* **68**, 303–310 (1986).
- Doucette, C. D., Schwab, D. J., Wingreen, N. S. & Rabinowitz, J. D. *Nature Chem. Biol.* **7**, 894–901 (2011).
- Inada, T., Kimata, K. & Aiba, H. *Genes Cells* **1**, 293–301 (1996).

MICROBIOLOGY

A weapon for bacterial warfare

The finding that bacteria use a sharp spike to deliver toxins into competing microorganisms, and that this mechanism co-evolved with a bacteriophage structure, presents a new vision of bacterial secretion systems. SEE LETTER P.350

ALAIN FILLOUX

Bacteria have evolved molecular machines to secrete toxins and proteins into their environment or into target cells. The distinctive features of these machines led microbiologists to classify them by type; the type VI secretion system¹, for example, has the unusual function of injecting toxins into bacterial prey, causing death². This system also shows unprecedented similarity to the

puncturing devices and injection machines of bacteriophages, viruses that infect bacteria³. Recognition of this remarkable likeness has led to notable discoveries, and that reported by Shneider *et al.*⁴ on page 350 of this issue might be viewed as the cherry on top*. The researchers have solved the structure of proteins belonging to the PAAR family and show how these proteins function in the

*This article and the paper under discussion⁴ were published online on 7 August 2013.

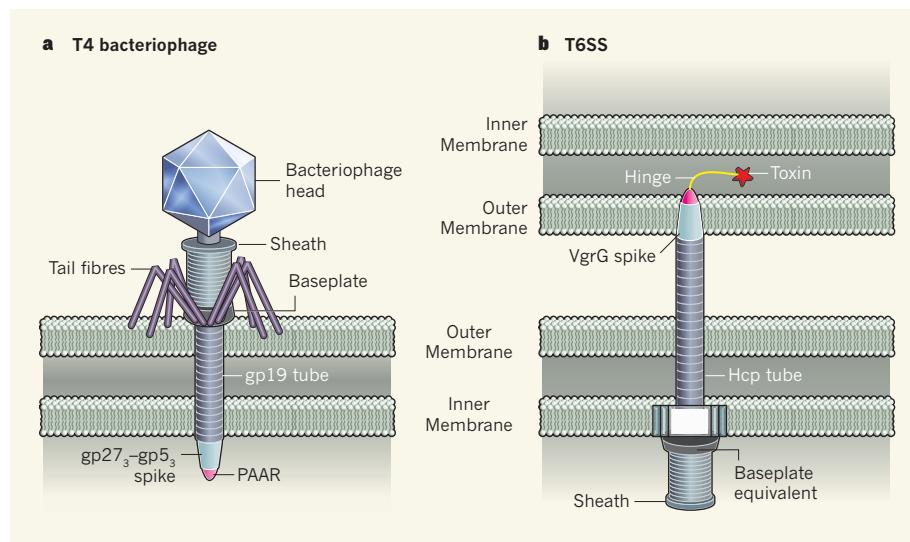


Figure 1 | Structures of the T4 bacteriophage and the bacterial type VI secretion system. Shneider *et al.*⁴ have established that small proteins belonging to the PAAR family sit on top of surprisingly similar protein structures that are found in the tail of T4 bacteriophage viruses (a) and the type IV secretion system (T6SS) of some bacteria (b). Both complexes contain a tube of hexameric protein rings (comprising gp19 in the bacteriophage and Hcp in T6SS) and a 'spike', made up of the VgrG protein in the T6SS and the gp27₃-gp5₃ protein complex in the T4 bacteriophage. Both tubes are surrounded by a sheath structure that contracts to push the spike through bacterial cell membranes; bacteriophages use this system to inject DNA into a cell they are infecting, whereas bacteria use the T6SS to inject proteins or toxins into other cells. Shneider and colleagues suggest that PAAR proteins associated with the T6SS act as adaptors that transport these proteins or toxins. The hinge (or a flexible domain) possibly connects the toxin to the PAAR.

type VI secretion system to deliver proteins to target cells — a breakthrough for our understanding of the assembly and evolution of this machine.

To follow this story, we need to appreciate how studies of bacteriophages have aided investigations of the action of the type VI secretion system (T6SS). Early database searches for genes with sequence similarity to those encoding T6SS proteins failed to provide leads as to the function of these genes, but insight came from the three-dimensional structure of two T6SS proteins, Hcp and VgrG. Hcp was shown to form hexameric rings that pile on top of each other to make a tube¹, and docking experiments³ showed structural similarities between Hcp and gp19, the protein that forms a tube in the tail of the T4 bacteriophage (Fig. 1). Furthermore, the structure of VgrG^{3,5} bears a striking resemblance to the gp27₃-gp5₃ protein complex in this bacteriophage⁶. This complex forms a needle-like structure ending in a rigid β-stranded helix; the needle functions as a puncturing device that sits on top of the gp19 tube. A contractile sheath structure pushes the tube and puncturing device through the envelope of bacterial cells, injecting the bacteriophage DNA into these target cells⁶. Electron microscopy studies^{7,8} have revealed that two proteins of T6SS form a similar sheath-like structure.

In possessing a sheath, a tube and a puncturing device, there is no doubt that the T6SS resembles the tail of a bacteriophage, although its directionality is inverted — it

pushes from inside the cell. Despite the fact that the systems clearly have different roles, with T6SS injecting proteins or toxins² and the bacteriophage tail injecting DNA, Shneider *et al.* sought to investigate just how far the structural similarities between the two go. The authors noted, from a reconstruction⁶ of the bacteriophage T4 tail made using cryo-electron microscopy, that a small but unidentifiable protein is located at the tip of the phage spike⁹. By analysing bacteriophage genomes and T6SS gene clusters, they found that both contain genes encoding small proteins of the PAAR family. They then used structural approaches³ to show that these proteins have a conical shape that fits perfectly on top of the gp27₃-gp5₃ (bacteriophage) and VgrG (T6SS) spikes. Moreover, the conical shape of the PAAR proteins makes the spikes sharper, emphasizing their role in puncturing bacterial cell membranes.

Most bacterial secretion systems are highly specific, such that a system from one bacterial species will not necessarily recognize and transport proteins from another, even if both species use a similar secretion system. Shneider and colleagues show that the T6SS is no exception, that PAAR proteins are essential for T6SS function, and that, although there is some flexibility, they cannot systematically be exchanged between T6SS systems of different origin. This led the authors to hypothesize that PAAR proteins could function as adaptors between the VgrG spike and the proteins or toxins to be transported (Fig. 1b).

A bioinformatics analysis showed that PAAR proteins are modular and that some have a carboxy-terminal extension with predicted enzymatic or toxin functions. This suggests that a PAAR protein sitting on the VgrG spike not only contributes to the piercing of the target cell, but also transports the toxin, which is bound to it. However, it is worth noting that some VgrG spikes already have functional C-terminal extensions — such as the actin crosslinking domain of the bacterium *Vibrio cholerae*¹⁰ — and may not need a PAAR adaptor.

This last part of Shneider and colleagues' study is more speculative and relies essentially on hypothesis. Indeed, the authors' interpretations could cast doubt on the role of PAAR as a sharp spike, if this part is hidden by a toxin or enzyme. Further structural evidence is required to assess whether adding a toxin or protein at its C terminus hinders the PAAR protein's puncturing activity. If it is shown not to, this will definitively reveal the T6SS to be a flexible and modular toxin-delivery device to which other domains could be added, multiplying the possible functions of the system. The authors' suggestion of such a mechanism means that the hunt for a broader panel of adaptors and toxins associated with the T6SS is now open.

The molecular details of this fascinating mechanism might distract from the question of why the T6SS has attracted so much attention. The T6SS is known as a system for killing other bacteria or subverting cells^{2,10} — a remarkable tool that can be used to build an optimal bacterial ecosystem, to colonize a host, or both, and which therefore has clear relevance to medicine and ecology. A secretion system per se has no function other than to transport proteins; the secreted proteins are the 'software' of the machines, and it is their activity that determines the physiological consequences of the secretion system. As such, the discovery of new T6SS effectors is likely to herald major breakthroughs. ■

Alain Filloux is at the MRC Centre for Molecular Bacteriology and Infection, Department of Life Sciences, Imperial College London, London SW7 2AZ, UK.
e-mail: a.filloux@imperial.ac.uk

- Mougous, J. D. *et al.* *Science* **312**, 1526–1530 (2006).
- Hood, R. D. *et al.* *Cell Host Microbe* **7**, 25–37 (2010).
- Leiman, P. G. *et al.* *Proc. Natl. Acad. Sci. USA* **106**, 4154–4159 (2009).
- Shneider, M. M. *et al.* *Nature* **500**, 350–353 (2013).
- Hachani, A. *et al.* *J. Biol. Chem.* **286**, 12317–12327 (2011).
- Leiman, P. G. *et al.* *Virol. J.* **7**, 355 (2010).
- Bönenmann, G., Pietrosik, A., Diemand, A., Zentgraf, H. & Mogk, A. *EMBO J.* **28**, 315–321 (2009).
- Lossi, N. S. *et al.* *J. Biol. Chem.* **288**, 7536–7548 (2013).
- Kostyuchenko, V. A. *et al.* *Nature Struct. Biol.* **10**, 688–693 (2003).
- Ma, A. T., McAuley, S., Pukatzki, S. & Mekalanos, J. J. *Cell Host Microbe* **5**, 234–243 (2009).

Climate extremes and the carbon cycle

Markus Reichstein¹, Michael Bahn², Philippe Ciais³, Dorothea Frank¹, Miguel D. Mahecha¹, Sonia I. Seneviratne⁴, Jakob Zscheischler^{1,4,5}, Christian Beer^{1,6}, Nina Buchmann⁴, David C. Frank^{7,8}, Dario Papale⁹, Anja Rammig¹⁰, Pete Smith¹¹, Kirsten Thonicke¹⁰, Marijn van der Velde¹², Sara Vicca¹³, Ariane Walz¹⁴ & Martin Wattenbach¹⁵

The terrestrial biosphere is a key component of the global carbon cycle and its carbon balance is strongly influenced by climate. Continuing environmental changes are thought to increase global terrestrial carbon uptake. But evidence is mounting that climate extremes such as droughts or storms can lead to a decrease in regional ecosystem carbon stocks and therefore have the potential to negate an expected increase in terrestrial carbon uptake. Here we explore the mechanisms and impacts of climate extremes on the terrestrial carbon cycle, and propose a pathway to improve our understanding of present and future impacts of climate extremes on the terrestrial carbon budget.

For the past five decades terrestrial ecosystems have been absorbing 25–30% of anthropogenic carbon dioxide (CO₂) emissions¹, with much of this uptake occurring via carbon accumulation in forest biomass and soils². Proposed mechanisms for this net carbon sink due to enhanced vegetation growth include CO₂ and nitrogen fertilization, and gradually increasing growing season length in northern regions. Overall, this terrestrial sink mitigates the anthropogenic increase of atmospheric CO₂ levels, and provides a negative feedback in the climate/carbon-cycle system³. It is essential to investigate to what extent, for how long and in which ecosystems this net CO₂ absorption and negative feedback will continue.

To address these questions, coupled carbon–climate model experiments have been set up⁴, the most recent being the Coupled Model Intercomparison Project (CMIP5), whose results are used in the current Intergovernmental Panel on Climate Change (IPCC) assessment. In the CMIP5 comparison, state-of-the-art Earth system models incorporate a mechanistic description of the carbon cycle coupled with climate. In all of these models, the combined effect of CO₂, climate change and (in fewer models) nitrogen deposition leads to increased vegetation productivity, fostering enhanced carbon sinks in temperate and boreal regions and the above-mentioned negative feedback on climate change. Future projections of ecosystem responses and thus feedback strength are, however, highly uncertain^{3,4}. Recent studies indicate that the occurrence of extreme events, for instance heatwaves, droughts or storms, and the associated disturbances can partially offset carbon sinks or even cause net losses in carbon stocks, thereby releasing CO₂ to the atmosphere^{5–8}. Because extreme events can trigger immediate and time-lagged responses of ecosystems, such as mortality, fires or insect infestations^{9,10}, their effects on carbon fluxes and stocks are nonlinear. Thus, even a small shift in the frequency or severity of climate extremes could substantially reduce carbon sinks and may result in sizeable positive feedbacks to climate warming.

In this Perspective, we investigate the diverse impacts of climate extremes on the carbon cycle of terrestrial ecosystems. We start with a conceptual treatment of climate extremes from an impact point of view, analyse key ecosystem mechanisms triggered by climate extremes and make a first attempt to estimate the susceptibility of the carbon cycle in different ecosystem

types. Finally, based on a set of metrics on the magnitude of extremes calculated from multi-temporal Earth observation data and Earth system model simulations, we provide a first estimate of the relative magnitude of carbon-cycle deviations caused by climate extremes over the past 30 years.

We propose that climate extremes have the potential to significantly affect the carbon cycle regionally and globally. To obtain reliable estimates of the sign and magnitude of future carbon-cycle feedbacks, a better understanding and descriptions of both the occurrence of climate extremes themselves and the ecosystem carbon-cycle processes that are triggered by climate extremes need to be achieved. To this end, we advocate a new generation of ecosystem manipulation experiments dedicated to studying extreme events, targeted long-term carbon-cycle observations, and an emphasis on high-resolution climate and biosphere modelling.

Climate extremes and the biosphere

The study of climate and weather extremes has a long history in climatology and hydrology. This research has led to a commonly applied statistical framework for defining climate extremes^{11,12}. Yet such definitions of extremes, based on climate statistics alone, are not necessarily well suited for assessing the impact on ecosystems and their carbon cycle, as illustrated in the following thought experiment. Consider a year with precipitation of, say, 499 mm—this observation is regarded as a statistical extreme, if for the past 100 years annual precipitation has always been between 500 mm and 510 mm (although with considerable seasonal variation), despite the fact that this difference of 1–11 mm or 0.2–2% is negligible in terms of its ecosystem impact. Similarly in the real world, if monthly winter temperatures at high latitudes typically vary between –30 °C and –40 °C, a month with –25 °C again is extreme according to the climatological definition, but is far below any critical threshold for inducing an ecosystem response.

A specific definition of climate extremes relevant for terrestrial ecosystems is thus needed, where the extremeness in the expected response, not only in meteorological drivers, is considered. Accordingly, it has been suggested that an extreme climatic event should be defined as “an episode or occurrence in which a statistically rare or unusual climatic period alters ecosystem structure and/or functions well outside the bounds of

¹Max Planck Institute for Biogeochemistry, 07745 Jena, Germany. ²Institute of Ecology, University of Innsbruck, 6020 Innsbruck, Austria. ³IPSL—Laboratoire des Sciences du Climat et de l'Environnement CEA-CNRS-UVSQ, 91191 Gif sur Yvette, France. ⁴ETH Zurich, 8092 Zurich, Switzerland. ⁵Max Planck Institute for Intelligent Systems, 72076 Tübingen, Germany. ⁶Stockholm University, Department of Applied Environmental Science (ITM), and the Bert Bolin Centre for Climate Research, 10691 Stockholm, Sweden. ⁷Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland. ⁸Oeschger Centre for Climate Change Research, University of Bern, CH-3012 Bern, Switzerland. ⁹Department for Innovation in Biological, Agro-food and Forest Systems (DIBAF), University of Tuscia, 01100 Viterbo, Italy. ¹⁰Potsdam Institute for Climate Impact Research (PIK), 14773 Potsdam, Germany. ¹¹Institute of Biological and Environmental Sciences, University of Aberdeen, Aberdeen AB24 3UU, UK. ¹²International Institute for Applied Systems Analysis, Ecosystem Services and Management Program, A-2361 Laxenburg, Austria. ¹³Research Group of Plant and Vegetation Ecology, Department of Biology, University of Antwerp, B-2610 Wilrijk, Belgium. ¹⁴University of Potsdam, Institute of Earth and Environmental Science, 14476 Potsdam, Germany. ¹⁵Helmholtz Centre Potsdam, GFZ German Research Centre for Geosciences, Section 5.4 Hydrology, 14473 Potsdam, Germany.

what is considered typical or normal variability¹³. This definition is general enough to include meteorological constellations that are not extreme for a single variable but are extreme for a combination of variables (a multivariate extreme or compound event^{12,14}), such as a combined heatwave and drought, or a drought followed by extreme precipitation.

To emphasize the impact perspective further, we restate this definition of biosphere-relevant climate extremes as “conditions where an ecosystem function (such as carbon uptake) is higher or lower than a defined extreme percentile during a defined time period and over a certain area, traceable to single or multivariate anomalous meteorological variables”. As a consequence of this definition, the identification and detection of extremes is first focused on ecosystem diagnosis, and then requires the attribution of an extreme ecosystem impact to immediate and lagged effects of meteorological variables (also see “Carbon-cycle extremes in the satellite period” section).

There is clear evidence that extreme events not only affect the carbon cycle concurrently (for example, by reducing vegetation productivity or destroying carbon stocks during fire events), but can initiate lagged responses. For example, one year after an anomalously warm season, soil heterotrophic respiration was enhanced in a grassland, offsetting net ecosystem carbon uptake¹⁵; soil frost increased the sensitivity of heterotrophic respiration to summer drought in a forest¹⁶; and increased tree mortality was measured after severe droughts in many instances^{17–19}. Lagged and legacy effects of extreme events on ecosystem carbon cycling are poorly understood, and can potentially involve multiple synergistic and antagonistic mechanisms operating in parallel at different levels of organization and timescales, for which a hierarchical response framework has been suggested²⁰. These mechanisms include (1) diminished plant

resistance to abiotic stress (for example, via antioxidants, osmolytes or changes in membrane stability), pests and pathogens (for example, via altered secondary metabolites) and their effects on plant performance; (2) changes in the amount, quality and timing of litter and rhizodeposition; (3) effects on soil physical and chemical characteristics (soil organic matter fractions, aggregate stability, hydrophobicity²¹); and (4) shifts in plant, microbial and animal species composition (for example, increased fungi, because fungi are more drought-resistant than bacteria²²) and associated changes in carbon and nitrogen cycling, which feed back to (1)–(3)²³.

Future research should address such mechanisms in integrated studies at the ecosystem scale, and provide a mechanistic basis for projecting effects of extreme meteorological constellations on the carbon cycle. In addition to the above-mentioned mechanisms at ecosystem level, we also envisage lagged effects in societal and economic systems with substantial lagged responses in the carbon cycle, for example if increasing food prices caused by low yields, combined with risk of wind throw, were to encourage conversion of forests to croplands or grasslands.

Ecosystem-dependent processes and impacts

Climate extremes induce a suite of interconnected effects, all of which have the potential to alter the carbon balance of ecosystems profoundly on different timescales (Fig. 1). This is illustrated by the effect of heatwaves and dry spells: this type of climate extreme has a direct effect on CO₂ fluxes, because both photosynthesis and respiration respond to warmer temperature and soil moisture limitation. Moreover, these factors work synergistically at the leaf, ecosystem and the regional scales as follows. Drought leads to stomatal closure by plants, decreasing leaf transpiration and evaporative cooling, aggravating the effect of high air temperatures

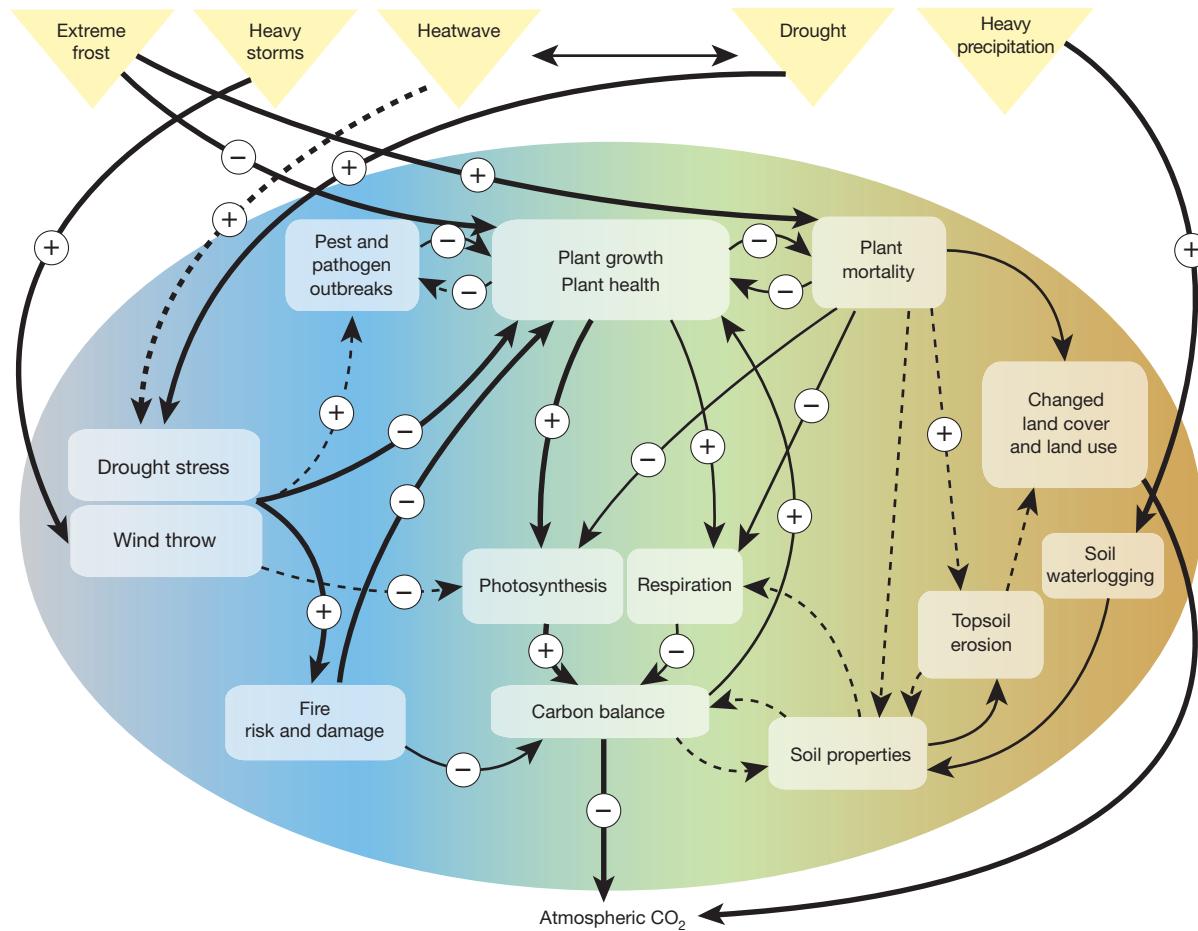


Figure 1 | Processes and feedbacks triggered by extreme climate events. The extreme events considered are droughts and heatwaves, heavy storms, heavy precipitation and extreme frost. Solid arrows show direct impacts;

dashed arrows show indirect impacts. The relative importance of the impact relationship is shown by arrow width (broader arrows are more important).

(and strong short-wave radiation)²⁴. Similarly, at the regional scale, soil moisture–temperature feedbacks can lead to a higher likelihood of heat-waves under dry soil conditions²⁵.

Moreover, soil drought, meaning more negative soil water potential and low soil hydraulic conductivity, usually leads to a higher vapour-pressure gradient between leaves and the atmosphere, causing a stress on the hydraulic system of plants that is further exacerbated by high temperatures. Consequently, high tension in the xylem can trigger embolism and partial failure of hydraulic transport in the stem, and can even be a contributory factor to mortality. This mechanism is currently regarded as a dominant cause of tree mortality under drought^{10,26}, interdependent with other postulated mechanisms related to the carbohydrate metabolism and insect infestations. In this respect, plant mortality can be a lagged effect of heat and drought that has impacts on the carbon balance for decades at least, and can lead to changing vegetation cover with associated feedbacks to local and regional hydrology and climate. Further effects of drought stress include an increase of fire risk and—as a delayed response—pathogen and pest outbreaks, the latter also being related to the sensitivity of ecosystems to heavy storms and wind-throw-related mortality in forests (Fig. 1).

Although the mechanisms triggered by different climate extremes can be described in conceptual terms as in the section above, their specific impact is highly dependent on ecosystem type. We summarize in Table 1, and the following section, specific anticipated effects of climate extremes for forests, peatlands, grasslands and croplands.

Forests and peatlands

Forests are characterized by the large biomass carbon stocks per square metre, which are vulnerable to wind throw, (ice-)storms, frost, drought, fire and pathogen or pest outbreaks. Moreover, given that trees take a long time to regrow, recovery times for forest biomass lost through extreme events are particularly long. Hence, the effects of climate extremes on the carbon balance in forests are both immediate and lagged, and potentially long-lasting. The impact of a climate extreme depends partially on events that have happened years before. Thus, assessing the net effect of climate extremes on forest carbon stocks requires a clearly defined time horizon during which immediate and lagged losses or gains can be assessed. In addition, any climate mitigation policy that relies on long-term carbon storage in forest biomass and the forest soil must guard against the likelihood of loss of accumulated carbon stocks in the face of future extreme events.

Although forests are potentially susceptible to all types of extreme event (Table 1), globally, drought is the most widespread factor affecting the carbon balance. For instance, during the European 2003 heatwave, precipitation (and soil moisture) deficit rather than temperature was the main factor reducing the water and carbon fluxes in the temperate and Mediterranean forest ecosystems^{27,28}. Severe and recurrent droughts have been identified as a major contributing factor in the recently accelerated rates of tree decline and mortality of forests^{17,29–31}. Beyond the most extensively studied mid-latitude forest belts, drought is also a relevant driver of the physiology and carbon cycling of the highly productive tropical

Table 1 | How forest, grasslands and croplands are affected by climate extremes

Land-cover type	Extremes	Key impact mechanisms	Examples of documented highly susceptible regions	Scientific understanding of future occurrence ¹²	Scientific understanding of carbon-cycle impact
Forest	Storms	<ul style="list-style-type: none"> Wind throw transforms carbon stock from living biomass to dry, dead wood Wind throw increases risk of fires and pathogen outbreaks 	The Amazon ³⁸ , North America ^{36,37} , central Europe ³⁵	Low	Medium
	Drought Heat	<ul style="list-style-type: none"> Water availability affects plant physiology, phenology and carbon allocation patterns Increased tree mortality, fire risk and susceptibility to pathogens Shifts in vegetation composition (impacts are large and delayed owing to the longevity of trees) Tree mortality has a large, fast impact on large carbon stocks in forests 	Central Europe ^{27,28} , western North America ³¹ , the Amazon ^{30,32}	Low to medium Medium to high	Low Low
	Fire		Western North America ⁷⁶ , southeastern Asia ⁷ , the Mediterranean ⁷⁷ , the circum-boreal areas ³⁹ , the Amazon ⁷⁸	Low	Low
	Ice storm and frost	<ul style="list-style-type: none"> Physical damage can include destruction of whole forest Xylem embolism and desiccation⁷⁹ 	China ⁸⁰ , North America ^{81,82}	Medium to high (for cold temperatures)	Low
Grasslands	Drought Heat	<ul style="list-style-type: none"> Species composition shifts (especially combined with additional pressure such as overgrazing) Degradation and desertification (especially combined with overgrazing) Erosion (combined with heavy precipitation or storms) 	North America ^{83,84} , Europe ⁸⁵ , central Asia ⁸⁶	Low to medium Medium to high	Medium Low
Croplands	Storms	<ul style="list-style-type: none"> Wind erosion and soil displacement with unclear consequences for the carbon cycle Direct crop damage 	China ⁸⁷ , North America ⁸⁸	Low	Low
	Heavy precipitation (including hail)	<ul style="list-style-type: none"> Erosion causing loss and displacement of soil and hence carbon Erosion affecting the soil's long-term productive capacity Crop damage or failure caused by hail and waterlogging of soils and subsequent anaerobic conditions Crop lodging, that is, the permanent displacement of cereal stems from the vertical Increase of pests and pathogens 	The tropics ⁸⁹ , North America ⁹⁰ , Australia ⁹¹ , the Mediterranean ⁹² , western Europe ⁵² , east Asia ⁹³	Medium to high (low for hail)	Low
	Drought and heat	<ul style="list-style-type: none"> Reduced growth or complete crop failure 	Europe ^{5,34,93} , North America ⁹⁴ , China ⁹⁵	Low to medium	Medium
	Extreme cold	<ul style="list-style-type: none"> Reduced growth Complete winter-crop failure, especially during spring frosts (combined with drought stress) 	North America ⁹⁶ , south Australia ⁹⁷ , Europe ⁹⁸	Medium to high	Low

forests²⁶. Amazonian forests were estimated to have lost 1.6 petagrams (10^{15} g) of carbon (Pg C) and 2.2 Pg C following the severe droughts of 2005 and 2010, respectively (refs 30 and 32). Soil conditions and rooting patterns play an important part in resilience and resistance³³.

Heavy storms and tropical cyclones regularly lead to severe forest damage, resulting in the loss of major limbs and foliage and widespread mortality⁶. Storms are considered to be the most important natural disturbance affecting European forests³⁴—the Lothar superstorm reduced the European standing forest biomass stocks in 1999 by about 16 teragrams (10^{12} g) of carbon (Tg C), corresponding to approximately 30% of the net biome production in Europe³⁵. Similarly, (sub-)tropical forests are vulnerable to wind-storm mortality and tropical-cyclone-driven mortality^{36,37}. A single squall line propagating across Amazonia in January 2005 caused widespread forest tree mortality, and threw to the ground the equivalent of 23% of the basin-wide mean annual biomass accumulation³⁸.

Around half of the average annual flux of 2.0 Pg C yr⁻¹ globally emitted by fires between 1997 and 2009 were from forest ecosystems, with 20% from deforestation and degradation fires in tropical forest, 16% from woodland fires, and 15% from (mostly extra-tropical) forest fires during the MODIS satellite era (2001–2009)³⁹. Fires and pest outbreaks can be facilitated by climate extremes in subtropical, Mediterranean and boreal forests and woodlands, but a direct link is not always easy to prove, because many factors, including direct human action, can trigger pests and fires. Nevertheless, climate anomalies associated with El Niño episodes have been shown to cause extreme fire events in tropical forests, affecting Amazon rainforest⁴⁰ and tropical southeast Asia, the latter contributing 66% of the atmospheric CO₂ growth rate anomaly during the 1997–1998 event⁴¹.

Peatlands have large carbon stocks, as forests do, but they are below ground and are mostly preserved by their high water table, which limits decomposition. Tropical peatlands contain approximately 90 Pg C worldwide and are particularly susceptible to drought and fires⁴². Northern peatlands contain 500 Pg C (ref. 43) and are susceptible to hydrological extremes, such as droughts and heavy rainfall, which in combination can lead to both losses of carbon by decomposition and immediate CO₂ release and by export of dissolved organic carbon, by which a quarter of net ecosystem production can be lost⁴⁴.

In summary, forest (and woodland) ecosystems are potentially susceptible to all climate extremes, with a plethora of important processes and indirect effects, as indicated in Fig. 1. With both large carbon stocks (standing biomass) and carbon fluxes being strongly affected by extremes, forest is the most sensitive biome to climate extremes. Yet many processes (for example, mortality) do not lead to immediate release of carbon to the atmosphere, but rather to a committed release via decomposition (Box 1, Fig. 2). If extremes are followed by fast regrowth and in each recovery cycle

carbon with increased residence times is generated (for example, by subsurface transport or charcoal formation), the whole process can be carbon neutral, or potentially can even lead to increased long-term sinks, though this has not yet been demonstrated.

Grasslands and savannas

Grasslands are susceptible to drought, whereas in contrast to forests, other extremes (for example, storms) play a smaller, if not negligible, part. Fire is often an intrinsic factor in grasslands and savannahs, and cannot be considered an extreme. It does, however, contribute to the suppression of trees in woody savannas (such as Miombo) and hence the suppression of build-up of high above-ground carbon stocks. Marcolla *et al.*⁴⁵ showed strong acclimation of a grassland to the interannual variability of climate, leading to a dampening of the interannual variability of the carbon balance in an Alpine grassland. Grasslands are also characterized by the high recovery potential of plant growth, as observed both in managed and in unmanaged grasslands^{46,47}, although the timing of the drought event may be a crucial factor in this regard⁴⁸. Overall, this high resilience distinguishes grasslands clearly from forests and explains why, on a global scale, grasslands often prevail in climates where extremely dry years can occur. However, once degradation feedbacks come into play, where drought triggers loss of vegetation and heavy rain causes subsequent erosion, more frequent extreme events may contribute to desertification of semi-arid to arid grassland, in particular when (over-)grazing is an additional pressure.

Croplands

There are at least three characteristics in which croplands differ from other ecosystems with respect to carbon-cycle responses to climate extremes. First, cropland systems are entirely managed, and management interventions can react on short timescales (for example, irrigation). Second, under annual cropping, the soil–vegetation system is reset regularly through harvest and agricultural management such as tillage, manure/residue management and irrigation. Third, type and duration of crop cover are highly variable, and the soil can be bare for an extended period of time. Consequently, the response to extremes is highly modulated by human intervention both immediately and over longer periods (for example, by changing cultivars or cultivation practice)^{49,50}.

Thus, the nature and extent of human interventions present one of the greatest uncertainties in assessing the impact of extremes on the carbon balance of croplands⁵⁰. Side effects will also very probably result from active interventions by farmers. Increased irrigation may enhance root biomass production, microbial activity and erosion rates, leading to increased or decreased soil organic carbon stocks. Regular harvesting and soil treatment makes long-term biological legacy effects of climate extremes more

BOX 1

Carbon processes triggered by climate extremes

Some extremes produce a direct biogeochemical signal in the atmosphere (Fig. 2, orange arrows), and may first be detected as regional anomalies in atmospheric CO₂ concentrations, analogous to a pollution plume. Globally, such signals are diluted by atmospheric mixing, but if large enough may still be detected as part of an annual growth anomaly.

In other cases, the effect of a climate extreme is essentially to alter the turnover rate of terrestrial carbon pools, leading to prolonged release of CO₂ to the atmosphere (see Fig. 1 for mechanisms) as a legacy from the extreme. For example, after a peak in mortality induced by wind throw or extreme drought, dead wood will take decades to decay. Such losses, often referred to as ‘committed CO₂ emissions’, may contribute to discrepancies between detected biosphere carbon-cycle anomalies and atmospheric CO₂ signals.

Although they may be less obvious, climate extremes may also trigger processes that decrease the turnover rate of some carbon pools and lead to additional long-term sequestration in these pools. A well-known example is the creation of charcoal during fire, which generally persists longer in soils than does the usual litter input⁹⁹.

Transport processes of particulate or dissolved organic carbon outside the ecosystem, either towards the subsurface or laterally into lake and river systems, have recently been shown to move relevant amounts of carbon⁵⁹. These processes can lead to both stabilization and destabilization of carbon, that is, they may decrease or increase turnover times. Indirect effects on biogeochemical cycling in lakes (such as stimulation of turnover) add to the complexity induced by the landscape-scale lateral processes, and the connection between ecosystems. The net effect of these transport processes on the carbon balance remains unclear.

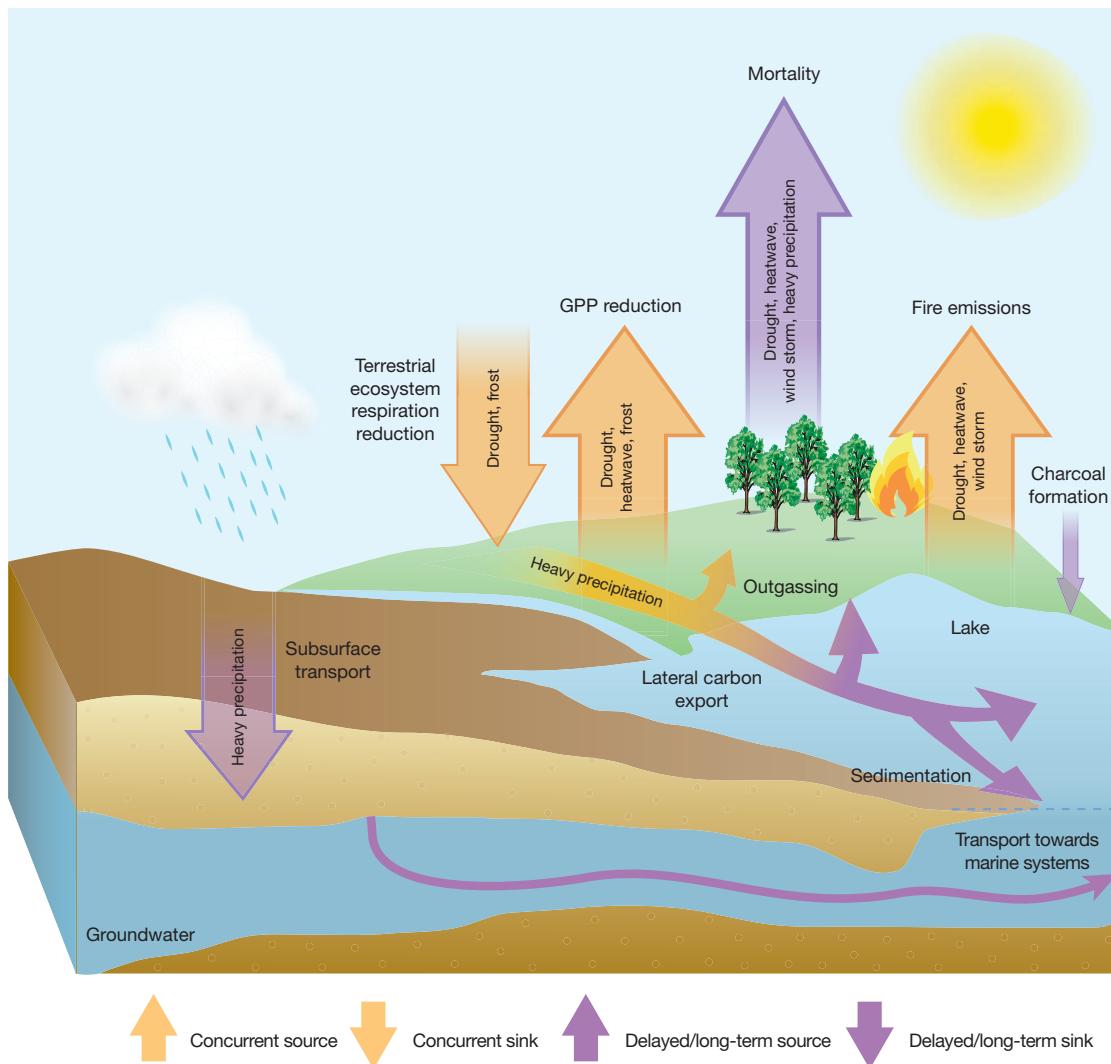


Figure 2 | Overview of how carbon flows may be triggered, or greatly altered, by extreme events. Emphasis is on the potential contrast between the concurrent and delayed signal in the atmosphere. Concurrent effects mean that the carbon signal can be found in the atmosphere while the climate extreme is occurring. Delayed and long-term signals occur either well after the extreme event has occurred, or are too small compared to the background short-term

variability to be immediately detectable. The concurrency and signal strengths involved with lateral transport probably depend on the transport distance, which is indicated by the colour gradient (from orange to purple). All concurrent fluxes can also be delayed, given the mechanisms and causal chains depicted in Fig. 1 and discussed in the main text. Sources and sinks are meant relative to conditions without extremes.

unlikely than in forests or grasslands, but the legacy effect due to the impacts of pathogen population dynamics cannot be excluded in crop lands. In addition, during a critical phase of its development, crop species can be vulnerable to a very specific stressor, which is otherwise unimportant during the rest of the year. For instance, rice pollen can become sterile if air temperature passes a 37°C threshold during the short pollination phase in spring⁵¹. In general, the impact of an extreme event is a crop-specific function of the timing of the extreme, in relation to the sensitivity of the plant during its growth stage⁵². Note also that crop yield can be decoupled from the carbon balance in the face of extreme climate conditions. For instance, a very high carbon uptake caused by an exceptionally warm winter did not induce high yield at a wheat site in Belgium, because of unfavourable weather conditions during the grain maturation stage⁵³.

Carbon-cycle extremes in the satellite period

Satellites enable us to evaluate the state of land vegetation. For instance, the fraction of absorbed photosynthetically active radiation (FAPAR) often serves as a spatiotemporal indicator for vegetation activity. In tandem with global networks of station measurements of land-atmosphere exchange

fluxes of CO_2 (refs 54 and 55) (that is, the FLUXNET initiative), modern machine-learning methods allow us to translate FAPAR into robust estimators for gross primary productivity (GPP). Today these continuous global data streams form a natural basis on which to examine the past three decades and to quantify the impacts of climate extremes on the carbon cycle across the world's ecosystems.

Zscheischler *et al.*⁵⁶ followed this impact-oriented search strategy and inventoried three decades of extremes in FAPAR anomalies with an emphasis on large spatiotemporally contiguous events. Using a 10% threshold to define FAPAR extremes, they estimated that the associated decrease in global GPP amounts on average to 2.7 Pg C yr^{-1} during the period 1982–2011 (using the fifth and first percentiles, the GPP decreases are 1.9 Pg C yr^{-1} and 0.7 Pg C yr^{-1} , respectively). Regionally, the most pronounced reduction of GPP (64 Tg C yr^{-1}) was observed in Northern Asia (following the IPCC nomenclature for regions¹²). Further, the most vulnerable regions to extremes are eastern Africa (55 Tg C yr^{-1}), eastern Asia (53 Tg C yr^{-1}), northeastern Brazil (53 Tg C yr^{-1}) and central North America (47 Tg C yr^{-1}). An analogous analysis of the CMIP5 model runs⁴ over the same period results in a global decrease of GPP due to extremes of $9.60 \text{ Pg C yr}^{-1}$ at the tenth percentile (5%: $6.56 \text{ Pg C yr}^{-1}$; 1%:

$2.28 \text{ Pg C yr}^{-1}$) averaged over all model runs. This higher sensitivity of GPP from models compared to Earth observation data could arise from model bias and (in tropical regions) from the difficulty of obtaining good fAPAR data from satellite observations and a consequent underestimation of GPP interannual variability in the satellite–FLUXNET data produced⁵⁷.

Attributing the observed fAPAR extremes over the past 30 years to anomalies in either temperature and/or water availability, or to fire events yields the strongest association with drought⁵⁶. Globally, from the hundred largest negative fAPAR extremes identified from satellite measurements between 1982 and 2011, 56 events are explainable by water scarcity, 14 by extreme high temperatures and 10 by exceptionally large fires (note, however, that of the hundred largest events only 43 fall in the time frame where state-of-the-art fire data are available, that is, 1997–2010). Thirty-five negative fAPAR extreme events could not be attributed to a particular driver. A regional exploration reveals systematic associations of negative fAPAR extremes with high water deficits in Africa, India and south-western Russia (Fig. 3a). Associations with both high temperatures and water deficit are found over Australia, North America and South America. Some of the remaining fAPAR extreme events, not attributable

to temperature or drought, might be caused by large-scale wind throw, by biotic events such as pest outbreaks or by a complex response to multivariate extremes (extremes not manifested in individual variables but in their uncommon coincidence). The total impact of extreme, spatiotemporally contiguous fAPAR anomalies on GPP is most strongly influenced by the spatial extent at which ecosystems have been affected by regionally extreme climate conditions (Fig. 3b). In contrast, the maximum intensity and the duration of each regional GPP anomaly are of secondary relevance.

It has been proposed that the size distribution of disturbance events in terrestrial ecosystems scales with a power law, like multiple other processes that are governed by phenomena of self-organized criticality⁵⁸. Scrutinizing the distributions of spatiotemporally integrated GPP extremes supports this contention. Zscheischler *et al.*⁵⁶ found that the magnitude of GPP extremes (fifth percentile) follows a power law with an exponent of α lying in the remarkably narrow range of 1.74 ± 0.02 across all vegetated continents (Fig. 3c). These values of α are consistent with previous estimates for the spatial extent of disturbance events in tropical and subtropical forests⁵⁸. On the one hand, the convergence of these macro properties points towards a coherent description and scaling of extremes in the global carbon cycle. On the other hand, we can also

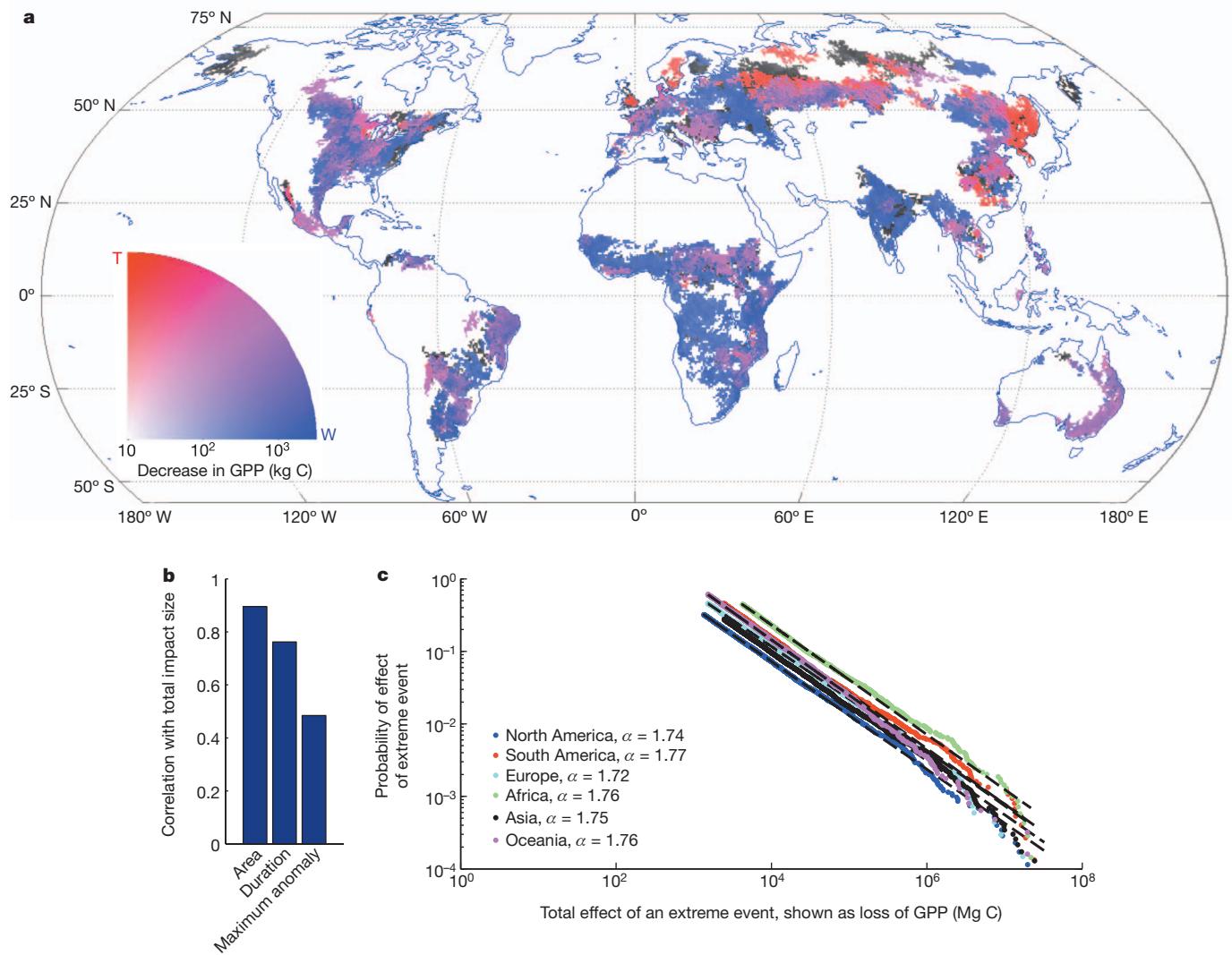


Figure 3 | Global impact of extreme events on the carbon cycle. **a**, Global distribution of extreme events impacting the terrestrial carbon cycle, defined as contiguous regions of extreme anomalies of fAPAR (lower first percentile) during the period 1982–2011 following the nomenclature of ref. 56. The hundred largest events on each continent are shown, along with whether they can be associated with water scarcity (blue, W), extreme high temperatures (red, T), both (pink) or neither (grey). The colour reflects the intensity of the

extreme event in terms of integrated loss of GPP, as indicated in the inset. **b**, Correlation of spatiotemporally integrated event impact size with maximum spatial extent ('Area'), duration ('Duration') and maximum intensity ('Maximum anomaly'). **c**, Size distribution of spatiotemporally integrated event impacts, following a power law with similar scaling exponents across all continents. α is the scaling exponent.

differentiate between different continents: Africa and South America are most likely to be hit by an extreme GPP impact of a given magnitude (Fig. 3c).

Spatiotemporal context and quantification

Detecting spatially and temporally contiguous extreme impacts in the carbon cycle and attributing them to climate extremes is naturally limited by data availability. Especially, the quantification and attribution of effects of extreme events on the carbon balance strongly depends on whether the considered time frame also captures lagged responses, and other carry-over effects as well as ecosystem recovery processes (see also Box 1 and Figs 1 and 2). Clearly, the spatial scale of integration affects the quantification of impacts of climate extremes on the carbon cycle, for example, when the climate extreme causes a transfer of carbon from one ecosystem to another. Inland waters are only recently considered as important sources or sinks of carbon following extreme events.

Carbon stored in lakes and rivers originates from soils and wetlands, with a smaller contribution from autochthonous net ecosystem production. In particular, heavy precipitation events may cause substantial lateral transfers of particulate and dissolved organic carbon from terrestrial to aquatic ecosystems, where it may be partly respiration and partly buried⁹, leading to an increased carbon sequestration. Aquatic systems are estimated to be the second most important ecosystem in terms of carbon sequestration, ranking above croplands and grassland in Europe⁵⁹. Further, erosion in general transports carbon from one system to a neighbouring one, where the imported carbon may be accumulated and turned over, and may potentially also cause carbon mobilization of autochthonous carbon pools⁶⁰.

Coming to grips with carbon-cycle extremes

As shown above, climate extremes trigger anomalous pulses in the carbon cycle, which can temporarily offset carbon sinks (see references in Table 1) and create CO₂ concentration signals detectable by large-scale atmospheric observations if the spatial extent of the affected region is large enough⁶¹. The analysis by Zscheischler *et al.* (ref. 56 and the above section) indicates that anomalies in vegetation CO₂ uptake induced by climate extremes add up to an average signal of global relevance, that is, of similar magnitude to that of the terrestrial carbon sink¹. Yet, on a global scale, the global growth rate of atmospheric CO₂ indicates that the land and oceans have continued to take up CO₂ with nearly the same relative strength as in the past⁶². This suggests that negative impacts of climate extremes on the global terrestrial carbon sink have so far not been increasing or decreasing disproportionately, so that on a larger scale the net biome production⁶³ (that is, the large-scale carbon balance including uptake, respiration, fire emissions, leaching, lateral transport and harvests) remains in the long term a constant fraction of anthropogenic CO₂ emissions. The terrestrial biosphere is the largest contributor to the year-to-year variability of the atmospheric CO₂ growth rate, but its relation to climate variability is not fully understood, leaving interannual residuals ranging between -2 Pg C yr^{-1} and $+2 \text{ Pg C yr}^{-1}$ when trying to close the global carbon balance¹. The effects of climate extremes on the carbon cycle have the potential to explain part of this residual variation. However, simple relationships between CO₂ growth rates and climate extremes are not necessarily expected, because of the above-mentioned lag and legacy effects involving ‘committed’ CO₂ emissions—regionally distinct constellations of climate variables causing carbon extremes, and differences in the response of ecosystems modulated by management.

Overall, the most pressing question is to what extent the coupling between increasing climate extremes¹² and induced CO₂ losses to the atmosphere might offset or even outweigh ecosystem carbon uptake arising from gradually increasing CO₂ concentration, prolonged extratropical growing season length and nitrogen deposition. Available observations and state-of-the-art modelling efforts are insufficient to provide definitive answers. Observational records of climate impacts on ecosystems and the carbon cycle are often too short and not widespread enough to provide sufficient context to interpret and mechanistically understand rare events. Continued simultaneous site-level observations of climate variables,

together with carbon and water cycles, such as those being carried out in a global observation network (FLUXNET⁶⁴) and global satellite-based observation of atmosphere and biosphere states, remain pivotal in this context. However, these should be more strongly complemented with longer time series from archived impacts, for instance in tree rings⁶⁵ or sediments⁶⁶.

Moreover, we need targeted assessments in regions where impacts of climate extremes have occurred, emphasizing the impact-oriented perspective proposed in this study. Nearly real-time information on the biosphere from satellite remote sensing and observation networks should enable rapid-response scientific campaigns to study after-effects and post-disturbance trajectories resulting from climate extremes. In particular, improved and repeated carbon state observations (such as biomass) from space would provide unprecedented information. Better mobilization of satellite imaging capabilities (that is, pointing satellites to an area affected by extremes, fast access to satellite data as for tsunami alerts⁶⁷) and online data processing and distribution of flux tower and atmospheric station data⁶⁸ will also offer the research community new tools with which to study carbon extremes and provide information needed for policy-making. Given the importance of drought, we strongly advocate stronger integration of carbon-related and hydrological observations and modelling efforts. More extensive high-quality soil moisture networks are critically important to better sample droughts in the affected regions⁶⁹.

Apart from long-term and ‘fast-response’ observations, it is also particularly important that future event-based manipulation experiments specifically test for thresholds and tipping points^{70,71} of ecologically relevant processes that can trigger long-lasting changes. In current manipulation experiments, these nonlinearities are rarely tested for or identified⁷². To considerably improve our mechanistic understanding of ecosystem responses to extreme events, it is crucial to perform increasingly standardized sets of measurements that permit calculation of common metrics to help us deal with the enormous variation in manipulation intensities, timing, return frequencies and duration in a coherent manner⁷³. Comparison of heatwave experiments requires quantification of the manipulation as experienced by the biota in terms of temperature as well as physiological water availability. Furthermore, a comprehensive approach with relevant impact measurements made at different plant and ecosystem levels (for example, stress hormones, allocation patterns, changes in vegetation community, as well as responses of soil processes) would substantially increase the value of future experiments. In addition, targeted and rapid response studies taking advantage of naturally occurring extremes (which can be further modified experimentally) should be considered. Following our global analysis, where we find tropical regions strongly affected by climate extremes (Fig. 3), we suggest that future emphasis should be placed on the tropical biomes.

Despite remarkable progress over the past decade, climate models still do not realistically simulate most climate extremes, because their designs are optimized to represent the long-term transient changes in the Earth system⁷⁴. Correspondingly, there also remain large uncertainties in projections of changes in biosphere-relevant climate extremes¹² (Table 1). However, given that droughts and related hydrological processes seem to be the dominant regional trigger for carbon-cycle extremes, an improved representation of drought dynamics in climate models will be crucial. More realistic representation of atmospheric dynamics and particularly convective processes and cloud formation⁷⁵ as well as soil hydrology⁶⁹ will be required. Although recent emphasis has been on building Earth system models that integrate climate and biogeochemical cycles, higher-resolution modelling efforts and improved physics should provide the basis for better projections of future climate extremes and their consequences for Earth’s carbon cycle.

Knowledge gained from new observations and experiments oriented towards climate extremes will enable the most important internal ecosystem feedbacks to be included and appropriately parameterized in terrestrial biosphere models. Given that nutrient cycles modulate the response of carbon fluxes to extreme events, global carbon-cycle modellers should aim for multi-element modelling (carbon, nitrogen, phosphorus). As in

climate models, this will require the adoption of higher spatial process resolution in carbon-cycle models. Biological adaptation and the role of biodiversity also remain particularly unclear, and thus are hardly considered in state-of-the-art Earth system models. One would expect adaptation and functional biodiversity to dampen the effect of climate extremes on the carbon cycle, but this has not been sufficiently established to justify their inclusion in currently operational global models.

The same is true for land management effects and their relation with climate extremes. The lateral and vertical transport of carbon out of the ecosystem into the hydrosphere (groundwater, rivers, lakes) during precipitation extremes calls for more strongly landscape-oriented land surface models, where parts of the landscape are intimately coupled (see Box 1 and Fig. 2). We suggest that coupling between the carbon and water cycles requires urgent attention, including how the carbon state variables feed back to the water cycle (such as soil organic carbon effects on infiltration and water-holding capacity), thus affecting susceptibility to meteorological drought.

Received 23 October 2012; accepted 29 May 2013.

1. Le Quéré, C., Raupach, M. R., Canadell, J. G. & Marland, G. Trends in the sources and sinks of carbon dioxide. *Nature Geosci.* **2**, 831–836 (2009).
2. Pan, Y. et al. A large and persistent carbon sink in the world's forests. *Science* **333**, 988–993 (2011).
3. Friedlingstein, P. et al. Climate-carbon cycle feedback analysis: results from the C4MIP model intercomparison. *J. Clim.* **19**, 3337–3353 (2006).
4. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
- In this full overview of the CMIP5 modelling experiment, the results of which are publicly available, the model simulations from BCC-CSM1.1, CanESM2, CCSM4, GFDL-ESM2G, HADGEM2-CC, HadGEM2-ES, INM-CM4, IPSL-CM5A-LR/MR, MIROC-ESM-(CHEM), MPI-ESM-LR and NorESM1-M have been used.
5. Ciais, P. et al. Europe-wide reduction in primary productivity caused by the heat and drought in 2003. *Nature* **437**, 529–533 (2005).
- This integrated data and modelling analysis showed that the extreme European heatwave 2003 undid 3–5 years of mean carbon sequestration.
6. Zeng, H. C. et al. Impacts of tropical cyclones on U.S. forest tree mortality and carbon flux from 1851 to 2000. *Proc. Natl Acad. Sci. USA* **106**, 7888–7892 (2009).
7. Page, S. E. et al. The amount of carbon released from peat and forest fires in Indonesia during 1997. *Nature* **420**, 61–65 (2002).
8. Kurz, W. A. et al. Mountain pine beetle and forest carbon feedback to climate change. *Nature* **452**, 987–990 (2008).
9. Van Oost, K. et al. The impact of agricultural soil erosion on the global carbon cycle. *Science* **318**, 626–629 (2007).
- This paper estimated the net effect of erosion given sources and sinks induced by the transported material.
10. Anderegg, W. R. et al. The roles of hydraulic and carbon stress in a widespread climate-induced forest die-off. *Proc. Natl Acad. Sci. USA* **109**, 233–237 (2012).
- This is a comprehensive analysis of mechanisms causing drought-related tree mortality.
11. Coles, S. G. *An Introduction to Statistical Modeling of Extreme Values* (Springer, 2001).
12. Seneviratne, S. I. et al. in *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation. A Special Report of Working Groups I and II of the Intergovernmental Panel on Climate Change (IPCC SREX Report)* (eds Field, C. B. et al.) 109–230 (Cambridge Univ. Press, 2012).
- This report gives a full assessment of the observed past and projected future occurrence and severity of climate extremes.
13. Smith, M. D. An ecological perspective on extreme climatic events: a synthetic definition and framework to guide future research. *J. Ecol.* **99**, 656–663 (2011).
- This paper introduced the ecosystem-impact-oriented perspective on climate extremes.
14. Ghil, M. et al. Extreme events: dynamics, statistics and prediction. *Nonlinear Process. Geophys.* **18**, 295–350 (2011).
15. Arnone, J. A. III et al. Prolonged suppression of ecosystem carbon dioxide uptake after an anomalously warm year. *Nature* **455**, 383–386 (2008).
- This was the first experimental study showing evidence for year-long lag effects of temperature extremes without involvement of mortality.
16. Muhr, J., Borken, W. & Matzner, E. Effects of soil frost on soil respiration and its radiocarbon signature in a Norway spruce forest soil. *Glob. Change Biol.* **15**, 782–793 (2009).
17. Bréda, N., Huc, R., Granier, A. & Dreyer, E. Temperate forest trees and stands under severe drought: a review of ecophysiological responses, adaptation processes and long-term consequences. *Ann. For. Sci.* **63**, 625–644 (2006).
18. Bigler, C., Gavin, D. G., Gunning, C. & Veblen, T. T. Drought induces lagged tree mortality in a subalpine forest in the Rocky Mountains. *Oikos* **116**, 1983–1994 (2007).
19. Adams, H. D. et al. Climate-induced tree mortality: Earth system consequences. *Eos* **91**, 153–154 (2010).
20. Smith, M. D., Knapp, A. K. & Collins, S. L. A framework for assessing ecosystem dynamics in response to chronic resource alterations induced by global change. *Ecology* **90**, 3279–3289 (2009).
21. Goebel, M.-O., Bachmann, J., Reichstein, M., Janssens, I. A. & Guggenberger, G. Soil water repellency and its implications for organic matter decomposition—is there a link to extreme climatic events? *Glob. Change Biol.* **17**, 2640–2656 (2011).
- This paper reviews how soil hydrological properties change persistently in response to climate extremes.**
22. de Vries, F. T. et al. Land use alters the resistance and resilience of soil food webs to drought. *Nature Clim. Change* **2**, 276–280 (2012).
23. Heimann, M. & Reichstein, M. Terrestrial ecosystem carbon dynamics and climate feedbacks. *Nature* **451**, 289–292 (2008).
24. De Boeck, H. & Verbeeck, H. Drought-associated changes in climate and their relevance for ecosystem experiments and models. *Biogeosciences* **8**, 1121–1130 (2011).
25. Seneviratne, S. I., Lüthi, D., Litschi, M. & Schär, C. Land-atmosphere coupling and climate change in Europe. *Nature* **443**, 205–209 (2006).
26. Chatz, B. et al. Global convergence in the vulnerability of forests to drought. *Nature* **491**, 752–755 (2012).
27. Reichstein, M. et al. Reduction of ecosystem productivity and respiration during the European summer 2003 climate anomaly: a joint flux tower, remote sensing and modelling analysis. *Glob. Change Biol.* **13**, 634–651 (2007).
28. Granier, A. et al. Evidence for soil water control on carbon and water dynamics in European forests during the extremely dry year: 2003. *Agric. For. Meteorol.* **143**, 123–145 (2007).
29. Allen, C. D. et al. A global overview of drought and heat-induced tree mortality reveals emerging climate change risks for forests. *For. Ecol. Manage.* **259**, 660–684 (2010).
- This is a global assessment of tree mortality in the context of climate change.**
30. Phillips, O. L. et al. Drought sensitivity of the Amazon rainforest. *Science* **323**, 1344–1347 (2009).
31. van Mantgem, P. J. et al. Widespread increase of tree mortality rates in the western United States. *Science* **323**, 521–524 (2009).
32. Lewis, S. L., Brando, P. M., Phillips, O. L., van der Heijden, G. M. F. & Nepstad, D. The 2010 Amazon drought. *Science* **331**, 554 (2011).
33. Nepstad, D. C., Tohver, I. M., Ray, D., Moutinho, P. & Cardinot, G. Mortality of large trees and lianas following experimental drought in an Amazon forest. *Ecology* **88**, 2259–2269 (2007).
34. Fuhrer, J. et al. Climate risks and their impact on agriculture and forests in Switzerland. *Clim. Change* **79**, 79–102 (2006).
35. Lindroth, A. et al. Storms can cause Europe-wide reduction in forest carbon sink. *Glob. Change Biol.* **15**, 346–355 (2009).
36. Chambers, J. Q. et al. Hurricane Katrina's carbon footprint on U.S. Gulf Coast forests. *Science* **318**, 1107 (2007).
37. Negrón-Jáurez, R., Baker, D. B., Zeng, H., Henkel, T. K. & Chambers, J. Q. Assessing hurricane-induced tree mortality in U.S. Gulf Coast forest ecosystems. *J. Geophys. Res.* **115**, G04030 (2010).
38. Negrón-Jáurez, R. I. et al. Widespread Amazon forest tree mortality from a single cross-basin squall line event. *Geophys. Res. Lett.* **37**, L16701 (2010).
39. van der Werf, G. R. et al. Global fire emissions and the contribution of deforestation, savanna, forest, agricultural, and peat fires (1997–2009). *Atmos. Chem. Phys.* **10**, 11707–11735 (2010).
40. Alencar, A., Nepstad, D. & Vera Diaz, M. C. Forest understory fire in the Brazilian Amazon in ENSO and non-ENSO years: area burned and committed carbon emissions. *Earth Interact.* **10**, 1–17 (2006).
41. van der Werf, G. R. et al. Continental-scale partitioning of fire emissions during the 1997 to 2001 El Niño/La Niña period. *Science* **303**, 73–76 (2004).
42. Page, S. E., Riley, J. O. & Banks, C. J. Global and regional importance of the tropical peatland carbon pool. *Glob. Change Biol.* **17**, 798–818 (2011).
43. Yu, Z. Northern peatland carbon stocks and dynamics: a review. *Biogeosciences* **9**, 4071–4085 (2012).
44. Dinsmore, K. J. et al. Role of the aquatic pathway in the carbon and greenhouse gas budgets of a peatland catchment. *Glob. Change Biol.* **16**, 2750–2762 (2010).
45. Marcolla, B. et al. Climatic controls and ecosystem responses drive the inter-annual variability of the net ecosystem exchange of an alpine meadow. *Agric. For. Meteorol.* **151**, 1233–1243 (2011).
46. Zavalloni, C. et al. Does a warmer climate with frequent mild water shortages protect grassland communities against a prolonged drought? *Plant Soil* **308**, 119–130 (2008).
47. Gilgen, A. K. & Buchmann, N. Response of temperate grasslands at different altitudes to simulated summer drought differed but scaled with annual precipitation. *Biogeosciences* **6**, 2525–2539 (2009).
48. De Boeck, H. J., Dreesen, F. E., Janssens, I. A. & Nijs, I. Whole-system responses of experimental plant communities to climate extremes imposed in different seasons. *New Phytol.* **189**, 806–817 (2011).
49. Lobell, D. B., Sibley, A. & Ortiz-Monasterio, J. I. Extreme heat effects on wheat senescence in India. *Nature Clim. Change* **2**, 186–189 (2012).
50. Porter, J. R. & Semenov, M. A. Crop responses to climatic variation. *Phil. Trans. R. Soc. Lond. B* **360**, 2021–2035 (2005).
51. Matsui, T., Narmuco, O. S., Ziska, L. H. & Horie, T. Effects of high temperature and CO₂ concentration on spikelet sterility in indica rice. *Field Crops Res.* **51**, 213–219 (1997).
52. van der Velde, M., Tubiello, F. N., Vrieling, A. & Bouraoui, F. Impacts of extreme weather on wheat and maize in France: evaluating regional crop simulations against observed data. *Clim. Change* **113**, 751–765 (2012).

53. Aubinet, M. et al. Carbon sequestration by a crop over a 4-year sugar beet/winter wheat/seed potato/winter wheat rotation cycle. *Agric. For. Meteorol.* **149**, 407–418 (2009).

54. Jung, M. et al. Recent decline in the global land evapotranspiration trend due to limited moisture supply. *Nature* **467**, 951–954 (2010).

55. Schwalm, C. R. et al. Assimilation exceeds respiration sensitivity to drought: a FLUXNET synthesis. *Glob. Change Biol.* **16**, 657–670 (2010).

56. Zscheischler, J., Mahecha, M. D., Harmeling, S. & Reichstein, M. Detection and attribution of large spatiotemporal extreme events in Earth observation data. *Ecol. Inform.* **15**, 66–73 (2013). **This was the first analysis of spatiotemporally contiguous carbon-cycle extremes.**

57. Jung, M. et al. Global patterns of land-atmosphere fluxes of carbon dioxide, latent heat, and sensible heat derived from eddy covariance, satellite, and meteorological observations. *J. Geophys. Res.* **116**, G00J07 (2011). **This was the first completely data-driven analysis of joint global carbon, water and sensible heat fluxes.**

58. Fisher, J. I., Hurnt, G. C., Thomas, R. Q. & Chambers, J. Q. Clustered disturbances lead to bias in large-scale estimates based on forest sample plots. *Ecol. Lett.* **11**, 554–563 (2008).

59. Luyssaert, S. et al. The European land and inland water CO₂, CO, CH₄ and N₂O balance between 2001 and 2005. *Biogeosciences* **9**, 3357–3380 (2012).

60. Quinton, J. N., Govers, G., Van Oost, K. & Bardgett, R. D. The impact of agricultural soil erosion on biogeochemical cycling. *Nature Geosci.* **3**, 311–314 (2010).

61. Peters, W. et al. Seven years of recent European net terrestrial carbon dioxide exchange constrained by atmospheric observations. *Glob. Change Biol.* **16**, 1317–1337 (2010).

62. Ballantyne, A., Alden, C., Miller, J., Tans, P. & White, J. Increase in observed net carbon dioxide uptake by land and oceans during the past 50 years. *Nature* **488**, 70–72 (2012).

63. Schulze, E.-D., Wirth, C. & Heimann, M. Managing forests after Kyoto. *Science* **289**, 2058–2059 (2000).

64. Baldocchi, D. et al. The role of trace gas flux networks in the biogeosciences. *Eos* **93**, 217 (2012).

65. Babst, F. et al. 500 years of regional forest growth variability and links to climatic extreme events in Europe. *Environ. Res. Lett.* **7**, 045705 (2012).

66. Adrian, R. et al. Lakes as sentinels of climate change. *Limnol. Oceanogr.* **54**, 2283–2297 (2009).

67. Kanamori, H. Real-time seismology and earthquake damage mitigation. *Annu. Rev. Earth Planet. Sci.* **33**, 195–214 (2005).

68. Paris, J.-D. et al. In *EGU General Assembly Conf. Abstr.* 12397 (European Geophysical Union, 2012).

69. Seneviratne, S. I. et al. Investigating soil moisture-climate interactions in a changing climate: a review. *Earth Sci. Rev.* **99**, 125–161 (2010). **This is a comprehensive review of aspects of soil moisture in the climate system with emphasis on regional feedbacks and observational needs.**

70. Schaeffer, M., Carpenter, S., Foley, J. A., Folke, C. & Walker, B. Catastrophic shifts in ecosystems. *Nature* **413**, 591–596 (2001).

71. Jentsch, A., Kreyling, J. & Beierkuhnlein, C. A new generation of climate-change experiments: events, not trends. *Front. Ecol. Environ.* **5**, 365–374 (2007).

72. Beier, C. et al. Precipitation manipulation experiments—challenges and recommendations for the future. *Ecol. Lett.* **15**, 899–911 (2012).

73. Vicca, S. et al. Urgent need for a common metric to make precipitation manipulation experiments comparable. *New Phytol.* **195**, 518–522 (2012).

74. Schiermeier, Q. The real holes in climate science. *Nature* **463**, 284–287 (2010).

75. Stevens, B. & Feingold, G. Untangling aerosol effects on clouds and precipitation in a buffered system. *Nature* **461**, 607–613 (2009).

76. Westerling, A. L., Hidalgo, H. G., Cayan, D. R. & Swetnam, T. W. Warming and earlier spring increase western U.S. forest wildfire activity. *Science* **313**, 940–943 (2006).

77. Moreira, F. et al. Landscape-wildfire interactions in southern Europe: implications for landscape management. *J. Environ. Manage.* **92**, 2389–2402 (2011).

78. Aragão, L. E. O. C. et al. Spatial patterns and fire response of recent Amazonian drought. *Geophys. Res. Lett.* **34**, L07701 (2007).

79. Sperry, J. S. & Sullivan, J. E. M. Xylem embolism in response to freeze-thaw cycles and water-stress in ring-porous, diffuse-porous, and conifer species. *Plant Physiol.* **100**, 605–613 (1992).

80. Sun, Y., Gu, L., Dickinson, R. E. & Zhou, B. Forest greenness after the massive 2008 Chinese ice storm: integrated effects of natural processes and human intervention. *Environ. Res. Lett.* **7**, 035702 (2012).

81. Irland, L. C. Ice storms and forest impacts. *Sci. Total Environ.* **262**, 231–242 (2000).

82. Changnon, S. A. Characteristics of ice storms in the United States. *J. Appl. Meteorol.* **42**, 630–639 (2003).

83. Knapp, A. K. et al. Rainfall variability, carbon cycling, and plant species diversity in a mesic grassland. *Science* **298**, 2202–2205 (2002). **This is a classic paper showing the importance of precipitation variability (compared to mean precipitation) for net primary production.**

84. Craine, J. M. et al. Timing of climate variability and grassland productivity. *Proc. Natl Acad. Sci. USA* **109**, 3401–3405 (2012).

85. Smit, H., Metzger, M. & Ewert, F. Spatial distribution of grassland productivity and land use in Europe. *Agric. Syst.* **98**, 208–219 (2008).

86. Wang, Y. et al. The fluxes of CO₂ from grazed and fenced temperate steppe during two drought years on the Inner Mongolia Plateau, China. *Sci. Total Environ.* **410–411**, 182–190 (2011).

87. Wang, X., Oenema, O., Hoogmoed, W. B., Perdok, U. D. & Cai, D. Dust storm erosion and its impact on soil carbon and nitrogen losses in northern China. *Catena* **66**, 221–227 (2006).

88. Changnon, S. A. Impacts of 1997–98 El Niño-generated weather in the United States. *Bull. Am. Meteorol. Soc.* **80**, 1819–1827 (1999).

89. Zhao, Y., Wang, C., Wang, S. & Tibig, L. V. Impacts of present and future climate variability on agriculture and forestry in the humid and sub-humid tropics. *Clim. Change* **70**, 73–116 (2005).

90. Rosenzweig, C., Tubiello, F. N., Goldberg, R., Mills, E. & Bloomfield, J. Increased crop damage in the US from excess precipitation under climate change. *Glob. Environ. Change* **12**, 197–202 (2002).

91. Niall, S. & Walsh, K. The impact of climate change on hailstorms in southeastern Australia. *Int. J. Climatol.* **25**, 1933–1952 (2005).

92. Sánchez, J. L. & Fraile, R. Crop damage: the hail size factor. *J. Appl. Meteorol.* **35**, 1535–1541 (1996).

93. van der Velde, M., Wriedt, G. & Bouraoui, F. Estimating irrigation use and effects on maize yield during the 2003 heatwave in France. *Agric. Ecosyst. Environ.* **135**, 90–97 (2010).

94. Jung, B.-J. et al. Storm pulses and varying sources of hydrologic carbon export from a mountainous watershed. *J. Hydrol.* **440–441**, 90–101 (2012).

95. Chen, G. et al. Drought in the Southern United States over the 20th century: variability and its impacts on terrestrial ecosystem productivity and carbon storage. *Clim. Change* **114**, 379–397 (2012).

96. Simelton, E. Food self-sufficiency and natural hazards in China. *Food Security* **3**, 35–52 (2011).

97. Gu, L. et al. The 2007 eastern US spring freezes: increased cold damage in a warming world? *Bioscience* **58**, 253–262 (2008).

98. Zheng, B. Y., Chenu, K., Drecer, M. F. & Chapman, S. C. Breeding for the future: what are the potential impacts of future frost and heat events on sowing and flowering time requirements for Australian bread wheat (*Triticum aestivum*) varieties? *Glob. Change Biol.* **18**, 2899–2914 (2012).

99. Schmidt, M. W. I. et al. Persistence of soil organic matter as an ecosystem property. *Nature* **478**, 49–56 (2011).

Acknowledgements This work emerged from the CARBO-Extreme project, funded by the European Community's Seventh Framework Programme under grant agreement (FP7-ENV-2008-1-226701). We acknowledge the World Climate Research Programme's Working Group on Coupled Modelling, which is responsible for CMIP, and we thank the climate modelling groups (listed in the reference annotation to ref. 4 of this paper) for producing and making available their model output. For CMIP the US Department of Energy's Program for Climate Model Diagnosis and Intercomparison provides coordinating support and led development of software infrastructure in partnership with the Global Organization for Earth System Science Portals. J.Z. is part of the International Max Planck Research School for Global Biogeochemical Cycles. P.S. is a Royal Society–Wolfson Research Merit Award holder. S.V. is a postdoctoral research associate of the Fund for Scientific Research—Flanders. M.B. acknowledges the Austrian Science Fund (FWF).

Author Contributions M.R., M.B., P.C. and S.I.S. conceived and designed the manuscript. D.F., M.R. and M.B. created Fig. 1, M.R. created Fig. 2 and M.D.M. and J.Z. created Fig. 3, with associated analysis and interpretation. The other co-authors contributed to specific sections. M.R. wrote the manuscript with comments and contributions from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence should be addressed to M.R. (markus.reichstein@bge.jena.mpg.de).

Translating dosage compensation to trisomy 21

Jun Jiang¹, Yuanchun Jing¹, Gregory J. Cost², Jen-Chieh Chiang¹, Heather J. Kolpa¹, Allison M. Cotton³, Dawn M. Carone¹, Benjamin R. Carone¹, David A. Shivak², Dmitry Y. Guschin², Jocelyn R. Pearl², Edward J. Rebar², Meg Byron¹, Philip D. Gregory², Carolyn J. Brown³, Fyodor D. Urnov², Lisa L. Hall¹ & Jeanne B. Lawrence¹

Down's syndrome is a common disorder with enormous medical and social costs, caused by trisomy for chromosome 21. We tested the concept that gene imbalance across an extra chromosome can be *de facto* corrected by manipulating a single gene, *XIST* (the X-inactivation gene). Using genome editing with zinc finger nucleases, we inserted a large, inducible *XIST* transgene into the *DYRK1A* locus on chromosome 21, in Down's syndrome pluripotent stem cells. The *XIST* non-coding RNA coats chromosome 21 and triggers stable heterochromatin modifications, chromosome-wide transcriptional silencing and DNA methylation to form a 'chromosome 21 Barr body'. This provides a model to study human chromosome inactivation and creates a system to investigate genomic expression changes and cellular pathologies of trisomy 21, free from genetic and epigenetic noise. Notably, deficits in proliferation and neural rosette formation are rapidly reversed upon silencing one chromosome 21. Successful trisomy silencing *in vitro* also surmounts the major first step towards potential development of 'chromosome therapy'.

In the United States, about 1 in 300 live births carry a trisomy, half of which are for chromosome 21, which causes Down's syndrome. Down's syndrome is the leading genetic cause of intellectual disabilities and the millions of Down's syndrome patients across the world also face multiple other health issues, including congenital heart defects, haematopoietic disorders and early-onset Alzheimer's disease^{1,2}. Down's syndrome researchers have sought to define the genes on chromosome 21 most closely associated with Down's syndrome, but this has proven difficult due to high genetic complexity and phenotypic variability of Down's syndrome, confounded by normal variation between individuals^{1–3}. Despite progress with mouse models for Down's syndrome^{4,5}, there remains a need for better ways to understand the underlying cell and developmental pathology of human Down's syndrome, key to therapeutic design of any kind².

The last decade has seen great advances in strategies to correct single-gene defects of rare monogenic disorders, beginning with cells *in vitro* and in several cases advancing to *in vivo* and clinical trials⁶. In contrast, genetic correction of the over-dose of genes across a whole extra chromosome in trisomic cells has remained outside the realm of possibility. Our effort was motivated by the idea that functional correction of living trisomic cells may be feasible by inserting a single gene that can epigenetically silence a whole chromosome. An inducible system for such 'trisomy silencing' would have immediate translational relevance as a resource to investigate the cellular pathology and gene pathways affected in Down's syndrome, in a setting free from pervasive genetic or epigenetic variation that exists between individuals, sub-clones, or even isogenic cell isolates^{3,7,8}.

There is a natural mechanism to compensate the difference in dosage of X-linked gene copies between mammalian females (XX) and males (XY). This is driven by a large (~17 kilobases (kb) in human), non-coding RNA, *XIST*, which is produced exclusively from the inactive X chromosome⁹, and 'paints' (accumulates across) the interphase chromosome structure¹⁰. During early development, the *XIST* RNA induces numerous heterochromatin modifications and architectural changes which transcriptionally silence the inactive X chromosome and manifest cytologically

as a condensed Barr body (reviewed in refs 11, 12). There is evidence for some DNA sequence specificity to *XIST* function, as certain human genes escape X-inactivation¹³; however, autosomal chromatin has substantial capacity to be silenced^{14–16}. Understanding the full potential of an autosome to be silenced, however, requires examination under conditions that avoid creating a deleterious functional monosomy. The strategy pursued here meets that requirement and creates a tractable model to study the distinct biology of human chromosome inactivation.

As outlined in Fig. 1a, we set out to determine whether the human X-inactivation gene, *XIST*, could be inserted into one copy of chromosome 21, and enact a chromosome-wide change in its epigenetic state. We pursued zinc finger nuclease (ZFN)-driven targeted addition¹⁷ of an inducible *XIST* transgene to the gene-rich core of chromosome 21 in induced pluripotent stem (iPS) cells derived from a Down's syndrome patient. If accomplished, this milestone would provide a system to study Down's syndrome cell pathology and the first step towards a potential genetic/epigenetic approach to 'chromosome therapy'.

Insertion of *XIST* into a trisomic chromosome 21

Given its large size, neither the *XIST* gene nor its cDNA has previously been integrated in a targeted fashion. Thus, our first goal was to demonstrate that ZFNs could accurately insert the largest transgene to date, substantially larger than sequences commonly used for genome editing¹⁸. We first attempted this with a ~16-kb *XIST* transgene in a transformed cell line (HT1080), using established ZFNs to the *AAVS1* locus on chromosome 19 (ref. 19). This proved highly successful (our unpublished data). To extend this to chromosome 21, we engineered ZFNs to a 36-base pair (bp) sequence in intron 1 of the *DYRK1A* locus at chromosome 21q22 (as in Fig. 1b), and validated their robust activity (Supplementary Fig. 1a, b). We tested an even larger (~21 kb) construct containing near full-length *XIST* cDNA in HT1080 cells and demonstrated efficient, accurate addition to this gene-rich region (Supplementary Fig. 2a, b).

We next determined whether this was achievable in technically challenging iPS cells, which have unique therapeutic and developmental

¹Department of Cell and Developmental Biology, University of Massachusetts Medical School, 55 Lake Avenue North, Worcester, Massachusetts 01655, USA. ²Sangamo BioSciences, 501 Canal Boulevard, Richmond, California 94804, USA. ³Department of Medical Genetics, University of British Columbia, 2350 Health Sciences Mall, Vancouver, British Columbia V6T 1Z3, Canada.

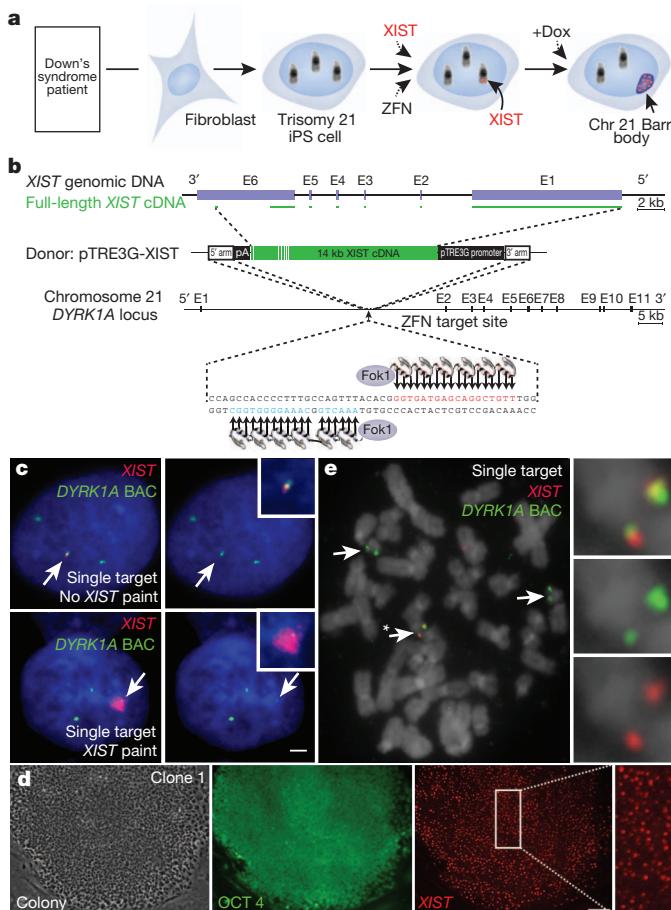


Figure 1 | Genome editing integrates *XIST* into chromosome 21 in trisomic iPS cells. **a**, Concept for translating dosage compensation to trisomy 21. **b**, *XIST* construct (19 kb): two homologous arms and 14-kb *XIST* cDNA with inducible pTRE3G promoter. **c**, DNA/RNA FISH in interphase Down's syndrome iPS cells shows that *XIST* overlaps one of three *DYRK1A* genes (left panels and insets) in a non-expressing cell (top, arrows), and a cell induced to express a large *XIST* RNA territory over the *DYRK1A* locus after 3 days in doxycycline (bottom, arrows). Right panels show green channel (*DYRK1A*) alone. Nuclear DNA is stained with 4',6-diamidino-2-phenylindole (DAPI, blue). Scale bar, 2 μ m. **d**, OCT4 immunostaining and *XIST* RNA FISH in a transgenic colony: highly consistent *XIST* expression throughout the colony. Scale bar, 100 μ m. **e**, Metaphase DNA FISH shows one targeted chromosome 21. *XIST* gene (asterisk and close-up) overlaps one of three *DYRK1A* genes (arrows).

potential to form various cell types, and thus would be important for any future *ex vivo* cellular therapy efforts. We used a male Down's syndrome iPS cell line²⁰ which we confirmed maintains pluripotency markers and trisomy 21. Although a constitutively transcribed transgene could be used, we engineered an inducible system to maximize utility for investigating Down's syndrome biology. In one step, we integrated both the doxycycline-controlled *XIST* transgene into chromosome 21 (Fig. 1b and Supplementary Fig. 3a) and a transgene carrying the doxycycline control component (rtTA) into the AAVS1 chromosome 19 safe harbour, disruption of which creates no known adverse effects¹⁹ (Supplementary Fig. 3b).

We analysed 245 colonies from the pooled transformants by interphase RNA/DNA fluorescence *in situ* hybridization (FISH) (Fig. 1c) to determine whether *XIST* was present and overlapped one of three *DYRK1A* alleles. Notably, 98.5% of *XIST*-positive colonies carried *XIST* at this location on chromosome 21, and also contained the rtTA/selection transgene (Supplementary Table 1). Efficiency was sufficiently high that, through modifications to editing conditions, we obtained a few sub-clones with *XIST* integrated into two or even

all three alleles of *DYRK1A* (see Methods and Supplementary Fig. 3c and Supplementary Table 1). Six independent sub-clones were chosen for further study based on: an *XIST* transgene on one of three chromosome 21 copies, pluripotent colony morphology, OCT4 (also called POU5F1) staining (Fig. 1d and Supplementary Fig. 4a), and formation of embryoid bodies. FISH to metaphase chromosomes (Fig. 1e and Supplementary Fig. 3d) and Southern blotting (Supplementary Fig. 1c–e) confirmed the gene addition accuracy, with 47 chromosomes, for all six clones. High-resolution cytogenetic banding and/or array comparative genomic hybridization (CGH) on selected clones showed no significant abnormalities other than full chromosome 21 trisomy (Supplementary Fig. 4c–e).

XIST RNA induces a chromosome 21 Barr body

In the panel of six independent genome-edited clones, we induced transgene expression and detected *XIST* RNA by FISH 3 days later. A localized *XIST* RNA 'territory' over one chromosome 21 (Fig. 1c) was seen in over 85% of cells in all six clones (Fig. 1d and Supplementary Fig. 4a, b). This mirrored the unique behaviour of endogenous *XIST* RNA which 'paints' the inactive X chromosome nuclear territory¹⁰.

The natural inactivated X chromosome forms a condensed Barr body which carries repressive histone marks¹¹. Similarly, 5 days after *XIST* induction, the edited chromosome 21 became markedly enriched in all heterochromatin marks examined, including H3K27me3, UbH2A and H4K20me in 90–100% of cells and, later, with macroH2A (Fig. 2a, b and Supplementary Fig. 5a). Supplementary Fig. 5b illustrates that H3K27me spreads across the whole metaphase chromosome 21. Moreover, chromosome 21 DNA in many nuclei became notably condensed, further evidence that we successfully generated a heterochromatic chromosome 21 Barr body (Fig. 2c).

Allele-specific silencing across chromosome 21

To measure overall transcription across the *XIST*-targeted chromosome 21, we used an approach that we developed to broadly assay heterogeneous nuclear RNA (hnRNA) expression and to distinguish inactive from active X chromosome¹⁶, on the basis of *in situ* hybridization to CoT-1 repeat RNA. This showed that the chromosome 21 *XIST* RNA territory was depleted for hnRNA detected by CoT-1 (Supplementary Fig. 5c), similar to the inactive X chromosome¹⁶.

We next used multi-colour RNA FISH to determine the presence of transcription foci at each allele for six specific chromosome 21 genes, an established approach that we earlier showed^{10,15} discriminates active versus silenced genes on inactive X chromosome. Without *XIST* expression, there are three bright transcription foci from each *DYRK1A* allele (Fig. 1c, top), but after *XIST* expression, the targeted allele becomes

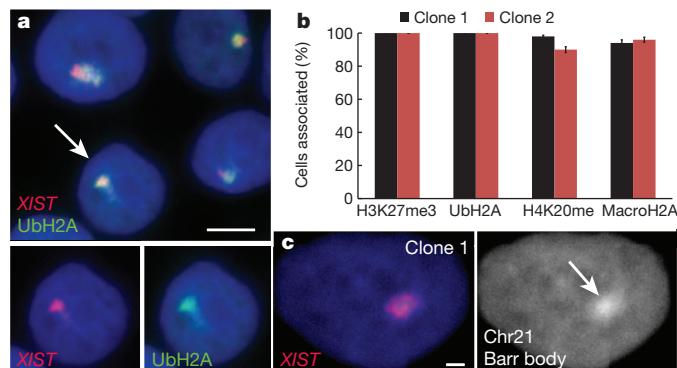


Figure 2 | *XIST* induces heterochromatin modifications and condensed chromosome 21 Barr body. **a**, *XIST* RNA recruits heterochromatic epigenetic marks (for example, UbH2A). Channels are separated for cell indicated with an arrow. Scale bar, 5 μ m. **b**, Percentage of *XIST* territories with heterochromatin marks. Mean \pm standard error, 100 nuclei in \sim 5 colonies. **c**, *XIST* RNA induces chromosome 21 Barr body visible by DAPI stain (arrow). Scale bar, 2 μ m.

weaker or undetectable, indicating repression of *DYRK1A* (Fig. 1c, bottom).

The *APP* gene on chromosome 21 encodes β-amyloid precursor protein; mutations in *APP* which cause accumulation of β-amyloid lead to early-onset familial Alzheimer's disease, and *APP* overexpression is linked to the Alzheimer's disease characteristic of Down's syndrome¹. Initially, three bright RNA transcription foci are apparent (Fig. 3a, top). Short-term *XIST* expression resulted in incomplete repression of the targeted allele (Fig. 3a, middle), which after 20 days was completely silenced, as shown in two independent clones (Fig. 3a, bottom, and Fig. 3b).

We examined four more loci, 3–21 megabases (Mb) from *XIST*: *ITSN1*, *USP25*, *CXADR* and *COL18A1*. Complete silencing of each allele on the edited chromosome 21 was seen in ~100% of cells accumulating *XIST* RNA (Fig. 3c, d and Supplementary Fig. 6a). Allele-specific silencing was further validated using single nucleotide polymorphism (SNP) analysis. PCR with reverse transcription (RT-PCR) products for eight known polymorphic sites (in four genes) were sequenced (ADAMTS1, *ETS2*, *TIAM1* and *HSPA13*) (Supplementary Fig. 6b, c). Interestingly, clones 2 and 3 showed an identical pattern of eight SNP alleles repressed, whereas clone 1 showed an alternative pattern of SNPs repressed. As summarized in Supplementary Fig. 6c, this chromosome-wide pattern allows extrapolation of the haplotype for each of the three chromosome 21 homologues, and indirectly identifies for each clone which chromosome 21 was silenced by an *XIST* transgene.

We also examined clones carrying *XIST* on two or all three copies of chromosome 21 and found that after 20 days in doxycycline, most or all cells lost *XIST* localization or expression, and the targeted chromosomes did not silence the *APP* gene (Supplementary Fig. 7a, b). Thus, there is *in vitro* selection and epigenetic adaptation to circumvent creating a functional monosomy or nullisomy, consistent with observations that monosomic cells do not persist in mosaic patients.

Genome-wide silencing and methylation

Having demonstrated allele-specific repression for the ten genes examined above, we extended this to genome-wide expression profiling. We treated three transgenic clones and the parental line with doxycycline for 3 weeks, and compared their transcriptomes to parallel cultures without *XIST* transcription, all in triplicate. Only on chromosome 21 is there overwhelming change, in all three clones (Fig. 4a), with ~95% of significantly expressed genes becoming repressed (Supplementary Table 2).

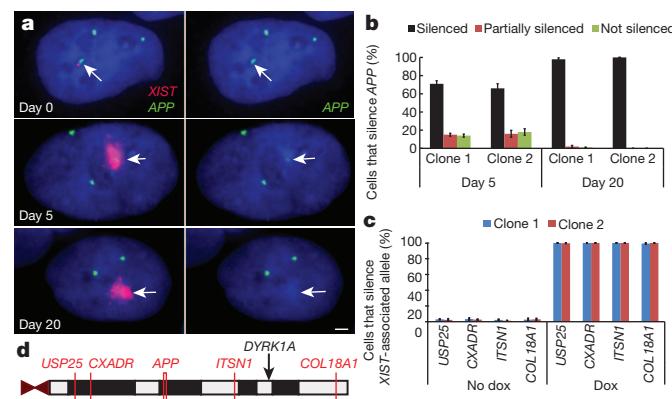


Figure 3 | *XIST* induces long-range silencing in targeted iPS cells. **a**, RNA FISH. APP RNA transcribes from three loci in uninduced cells (day 0), and is progressively silenced after induction (targeted chromosome 21, arrows). Scale bar, 2 μm. **b**, Quantification of APP silencing. Mean ± standard error, 100 nuclei. **c**, Silencing for four more chromosome-21-linked genes by RNA FISH. Mean ± standard error from 100 nuclei. **d**, Long-range silencing of chromosome 21 genes by *XIST* RNA. USP25 is ~21 Mb from the *XIST* integration site (black arrow).

Dosage compensation corrects chromosome 21 expression to near normal disomic levels, based on the change in total output of expressed genes per chromosome after *XIST* is induced. Because evidence suggests that many chromosome 21 genes are not increased the theoretical 1.5-fold in trisomy^{21,22}, we also directly compared trisomic to disomic cells. This provides a baseline for evaluating the degree to which chromosome 21 overexpression is corrected by *XIST*. After *XIST* induction, overall chromosome 21 expression is reduced by 20%, 15% and 19% for clones 1, 2 and 3, respectively; this mirrors very well the 22% reduction for disomic iPS cells that lack the third chromosome 21 altogether (Fig. 4a). This disomic line is representative, as a similar difference (21%) was seen for an isogenic disomic sub-clone that we isolated from the trisomic parental iPS cells (not shown). Individual genes repressed by *XIST* are distributed throughout chromosome 21, as do genes overexpressed in trisomic versus disomic cells (Fig. 4b). In addition, qRT-PCR confirmed repression for individually examined genes (Supplementary Fig. 7c). Clearly, *XIST* induces robust dosage compensation of most chromosome 21 genes overexpressed in trisomy.

Trisomy 21 may have an impact on genome-wide expression pathways, but differences attributable to trisomy 21 are confounded by genetic and epigenetic variability²¹. This inducible trisomy silencing system provides a new foothold into this important question. For example, even the three isogenic transgenic sub-clones show many expression differences (>1,000), but upon *XIST* induction, ~200 genes throughout the genome change in all three clones (but not the doxycycline-treated parental), most probably directly due to chromosome 21 trisomy. Therefore, 'trisomy correction in a dish' has promise as a means to identify genome-wide pathways perturbed by trisomy 21.

In addition to transcriptional silencing, X-inactivation is stabilized by hypermethylation of promoter CpG islands^{23,24}, which occurs late

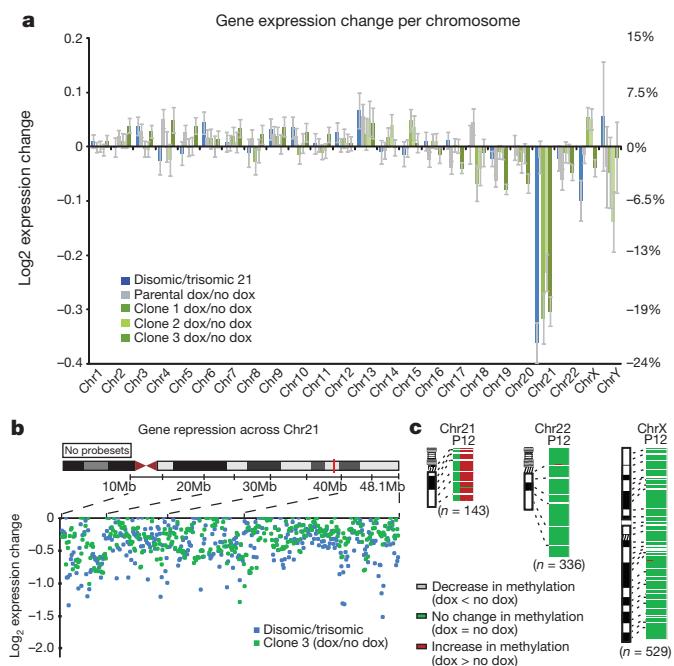


Figure 4 | Genomic expression and methylation reveal widespread silencing of chromosome 21. **a**, Microarray: expression difference for three transgenic clones in doxycycline (dox) versus no doxycycline, compared to disomic line versus trisomic parental line. Total change in gene expression ($n = 3$) per chromosome shows chromosome 21 'correction' near disomic levels, with only limited changes on other chromosomes. The right y axis is scaled for per cent gene expression change. Mean ± standard deviation, in triplicate. **b**, Distribution of individual gene repression across chromosome 21. **c**, Methylation of CpG island promoters. In treated clones, 97% of chromosome 21 genes increased by at least 5% (2-fold greater than average), compared to none in the parental line. P, parental line; 1, clone 1; 2, clone 2.

in the silencing process. Therefore, we also examined the promoter methylome in two genome-edited clones 3 weeks after *XIST* induction and found it largely unaltered, with one striking exception, genes on chromosome 21 (P value $<2.2 \times 10^{-16}$) (Fig. 4c). Here, 97% of CpG-island-containing genes exhibited a robust increase in promoter DNA methylation, within the range of that seen for the inactive X chromosome²⁴ (adjusted for active/inactive chromosomes; see Methods). This change swept the entire chromosome, with the interesting exception of a few genes that 'escape' methylation in both clones.

In summary, data from eight different approaches demonstrate impressive competence of most chromosome 21 genes to undergo epigenetic modification and silencing in response to an RNA that evolved to silence the X chromosome.

Phenotypic correction *in vitro*

Dosage compensation of chromosome imbalance presents a new paradigm, with opportunities to advance Down's syndrome research in multiple directions, including a new means to investigate human Down's syndrome cellular pathologies, which are largely unknown. Inducing trisomy silencing in parallel cultures of otherwise identical cells may reveal cellular pathologies due to trisomy 21, which could be obscured by differences between cell isolates. We examined cell proliferation and neural rosette formation to look for an impact on cell phenotype.

There is some evidence of proliferative impairment in Down's syndrome brains^{4,25}; however, we observed that this varied *in vitro* between our Down's syndrome fibroblast samples, and this would be highly sensitive to culture history. A clear answer emerged from comparing identical cell cultures, grown with or without doxycycline for 1 week. *XIST* induction in six independent transgenic sub-clones rapidly and consistently resulted in larger, more numerous and tightly packed colonies in just 7 days (Fig. 5a and Supplementary Fig. 8a), with 18–34% more cells (Fig. 5b). Doxycycline did not enhance growth of the parental Down's syndrome cells or sub-clone (Fig. 5b and Supplementary Fig. 8a). Thus, a proliferative impairment linked to chromosome 21 overexpression can be rapidly ameliorated by dosage compensation.

We next examined differentiation of targeted Down's syndrome iPS cells into neural progenitor cells. In 11–12 days after neural induction of already confluent cultures, all three *XIST*-expressing cultures began to form neural rosettes, and in 1–2 days were replete with

neural rosettes (Fig. 5c), a signature of neural progenitors (confirmed by PAX6 and SOX1 staining) (Supplementary Fig. 8b). Notably, even at day 14, parallel uninduced cultures remained devoid of rosettes (Fig. 5c). Uncorrected cultures required 4–5 more days in neural-induction media to fill with neural rosettes of similar size and number, which they did on day 17 (Fig. 5d and Supplementary Fig. 8d). There was no effect of doxycycline on neurogenesis in the parental line (Fig. 5c, d and Supplementary Fig. 8d). This marked delay in neural differentiation seems to be primarily independent of cell proliferation (Methods). Variability in the kinetics of neural differentiation between various iPS cell lines can obscure differences due to trisomy 21 (ref. 26). We circumvented this using parallel cultures and on-demand chromosome 21 silencing, which made clear these important phenotypic differences. This highlights the potential of this new experimental model to illuminate cellular pathologies directly attributable to chromosome 21 overexpression in iPS cells and their differentiated progeny.

Towards future applications

The Supplementary Information summarizes two significant points relevant to potential applications and therapeutic strategies. First, we show that heterochromatic silencing is stably maintained, even upon removal of doxycycline and *XIST* expression (Supplementary Fig. 9a, b), consistent with previous studies²³. Second, although not investigated extensively, we targeted *XIST* in non-immortalized fibroblasts from a female Down's syndrome patient, which generated many cells carrying *XIST* (and some heterochromatin marks) on chromosome 21 (Supplementary Fig. 9c, d). Finally, we note that our *XIST* transgene lacks X-chromosome 'counting' sequences, and thus is compatible with natural female X inactivation.

Discussion

We set out to bridge the basic biology of X-chromosome dosage compensation with the pathology of chromosomal dosage disorders, particularly Down's syndrome. In so doing, the present work yields advances that have an impact on three important areas: one basic and two translational.

Although not our primary focus here, a significant impact of this work is that we have created a tractable, inducible system to study human chromosome silencing. Importantly, unlike random integration into a diploid cell, silencing a trisomic autosome avoids selection against full autosomal silencing, and this demonstrated remarkably robust competence of chromosome 21 to be silenced. Thus, *XIST* RNA evolved for the X chromosome uses epigenome-wide mechanisms¹². The ability to insert a single *XIST* transgene in any locus provides a more powerful tool to study *XIST* function, and our effort also almost triples the size of transgenes that can be thus targeted for a host of other applications.

From a translational perspective, trisomy silencing has immediate impact as a new means to define the poorly understood cellular pathways deregulated in Down's syndrome, and creates the opportunity to derive and study various patient-compatible cell types potentially relevant to Down's syndrome therapeutics. Inducible trisomy silencing *in vitro* compares otherwise identical cultures, allowing greater discrimination of differences directly due to chromosome 21 overexpression distinct from genetic and epigenetic variation between transgenic sub-clones, or potentially even rare disomic sub-clones isolated from a trisomic population (refs 27, 28 and this study). *XIST* expression triggers not only chromosome 21 repression, but a defined effect on the genomic expression profile, and reverses deficits in cell proliferation and neural progenitors, which has implications for hypopcellularity in the Down's syndrome brain^{4,25}. This new approach can illuminate the cohort of genes and cognate pathways most consistently impacted in Down's syndrome, to inform the search for drugs that may rebalance those pathways and cell pathologies. This general strategy could be extended to study other chromosomal disorders, such as trisomy 13 and 18, often fatal in the first 1–2 years.

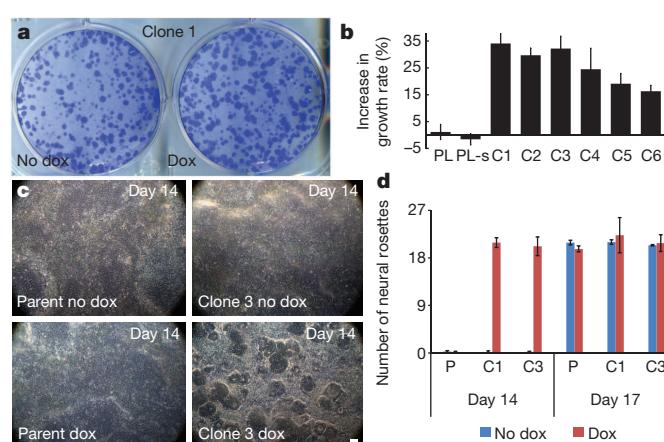


Figure 5 | 'Trisomy correction' affects cell proliferation and neurogenesis. **a**, One week of *XIST* expression resulted in larger, more numerous colonies (representative sample). **b**, Changes in cell number for parental line (PL), parental line subclone (PL-s), and six transgenic clones (C1–C6). Mean \pm s.e. ($n = 4$ –6). **c**, Corrected cultures formed neural rosettes by day 14; trisomic (parental and non-induced) cultures took longer (17 days). Scale bar, 100 μ m. **d**, Number of rosettes formed on day 14 and day 17. Mean \pm standard error, 10–12 random fields in triplicate. P, parental; C1, clone 1; C3, clone 3.

Finally, the more forward-looking implication of this work is to bring Down's syndrome into the realm of consideration for future gene therapy research. Although development of any clinical gene therapy is a multi-step process, any prospect requires that the first step, functional correction of the underlying genetic defect in living cells, is achievable. We have demonstrated that this step is no longer insurmountable for chromosomal imbalance in Down's syndrome. Our hope is that for individuals and families living with Down's syndrome, the proof-of-principle demonstrated here initiates multiple new avenues of translational relevance for the 50 years of advances in basic X-chromosome biology.

METHODS SUMMARY

ZFNs against the *DYRK1A* locus on chromosome 21 were designed and validated by established procedures¹⁸. These and previously identified ZFNs to chromosome 19 *AAVS1* (ref. 19) were used to deliver the *XIST* gene and rTA/puro to chromosome 21 and chromosome 19, respectively. All constructs were simultaneously electroporated into a Down's syndrome iPS cell line (DS1-iPS4) (G. Q. Daley, Children's Hospital Boston)²⁰. Over 100 clones were isolated and 6 chosen for more analysis. Silencing of the targeted chromosome 21 was demonstrated by eight different approaches as detailed in Methods (RNA microarray, DNA methylation array, RNA FISH to chromosome 21 genes, heterochromatin hallmarks, qRT-PCR, gene SNP analysis, Barr body formation, RNA FISH to hnRNA). Down's syndrome iPS cells were assessed for phenotypic differences (proliferation and neural rosette formation) before and after trisomy correction, as detailed in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 21 May 2012; accepted 18 June 2013.

Published online 17 July 2013.

1. Mégarbané, A. et al. The 50th anniversary of the discovery of trisomy 21: the past, present, and future of research and treatment of Down syndrome. *Genet. Med.* **11**, 611–616 (2009).
2. Gardiner, K. J. Molecular basis of pharmacotherapies for cognition in Down syndrome. *Trends Pharmacol. Sci.* **31**, 66–73 (2010).
3. Prandini, P. et al. Natural gene-expression variation in Down syndrome modulates the outcome of gene-dosage imbalance. *Am. J. Hum. Genet.* **81**, 252–263 (2007).
4. Haydar, T. F. & Reeves, R. H. Trisomy 21 and early brain development. *Trends Neurosci.* **35**, 81–91 (2012).
5. O'Doherty, A. et al. An aneuploid mouse strain carrying human chromosome 21 with Down syndrome phenotypes. *Science* **309**, 2033–2037 (2005).
6. Lee, B. & Davidson, B. L. Gene therapy grows into young adulthood: special review issue. *Hum. Mol. Genet.* **20**, R1 (2011).
7. Hall, L. L. et al. X-inactivation reveals epigenetic anomalies in most hESC but identifies sublines that initiate as expected. *J. Cell. Physiol.* **216**, 445–452 (2008).
8. Nazor, K. L. et al. Recurrent variations in DNA methylation in human pluripotent stem cells and their differentiated derivatives. *Cell Stem Cell* **10**, 620–634 (2012).
9. Brown, C. J. et al. The human *XIST* gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* **71**, 527–542 (1992).
10. Clemson, C. M., McNeil, J. A., Willard, H. F. & Lawrence, J. B. *XIST* RNA paints the inactive X chromosome at interphase: evidence for a novel RNA involved in nuclear/chromosome structure. *J. Cell Biol.* **132**, 259–275 (1996).
11. Heard, E. Delving into the diversity of facultative heterochromatin: the epigenetics of the inactive X chromosome. *Curr. Opin. Genet. Dev.* **15**, 482–489 (2005).
12. Hall, L. L. & Lawrence, J. B. *XIST* RNA and architecture of the inactive X chromosome: implications for the repeat genome. *Cold Spring Harb. Symp. Quant. Biol.* **75**, 345–356 (2010).
13. Carrel, L. & Willard, H. F. X-inactivation profile reveals extensive variability in X-linked gene expression in females. *Nature* **434**, 400–404 (2005).

14. Lee, J. T., Strauss, W. M., Dausman, J. A. & Jaenisch, R. A 450 kb transgene displays properties of the mammalian X-inactivation center. *Cell* **86**, 83–94 (1996).
15. Hall, L. L., Clemson, C. M., Byron, M., Wydner, K. & Lawrence, J. B. Unbalanced X;autosome translocations provide evidence for sequence specificity in the association of *XIST* RNA with chromatin. *Hum. Mol. Genet.* **11**, 3157–3165 (2002).
16. Hall, L. L. et al. An ectopic human *XIST* gene can induce chromosome inactivation in postdifferentiation human HT-1080 cells. *Proc. Natl. Acad. Sci. USA* **99**, 8677–8682 (2002).
17. Moehle, E. A. et al. Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. USA* **104**, 3055–3060 (2007).
18. Urnov, F. D., Rebar, E. J., Holmes, M. C., Zhang, H. S. & Gregory, P. D. Genome editing with engineered zinc finger nucleases. *Nature Rev. Genet.* **11**, 636–646 (2010).
19. DeKelver, R. C. et al. Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome. *Genome Res.* **20**, 1133–1142 (2010).
20. Park, I. H. et al. Disease-specific induced pluripotent stem cells. *Cell* **134**, 877–886 (2008).
21. At Yahya-Graison, E. et al. Classification of human chromosome 21 gene-expression variations in Down syndrome: impact on disease phenotypes. *Am. J. Hum. Genet.* **81**, 475–491 (2007).
22. Biancotti, J. C. et al. Human embryonic stem cells as models for aneuploid chromosomal syndromes. *Stem Cells* **28**, 1530–1540 (2010).
23. Csankovszki, G., Nagy, A. & Jaenisch, R. Synergism of *Xist* RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* **153**, 773–784 (2001).
24. Cotton, A. M. et al. Chromosome-wide DNA methylation analysis predicts human tissue-specific X inactivation. *Hum. Genet.* **130**, 187–201 (2011).
25. Guidi, S., Ciani, E., Bonasoni, P., Santini, D. & Bartesaghi, R. Widespread proliferation impairment and hypopcellularity in the cerebellum of fetuses with down syndrome. *Brain Pathol.* **21**, 361–373 (2011).
26. Shi, Y. et al. A human stem cell model of early Alzheimer's disease pathology in Down syndrome. *Sci. Transl. Med.* **4**, 124ra29 (2012).
27. Lavon, N. et al. Derivation of euploid human embryonic stem cells from aneuploid embryos. *Stem Cells* **26**, 1874–1882 (2008).
28. Li, L. B. et al. Trisomy correction in down syndrome induced pluripotent stem cells. *Cell Stem Cell* **11**, 615–619 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We appreciate recent initiatives by administrators of NIGMS and NIH to support more high-risk, high-impact research. Research began with support from GM053234 to J.B.L. for basic X chromosome research, and was made fully possible by GM085548 and GM096400 RC4 to J.B.L. C.J.B. and A.M.C. were supported by CIHR (MOP-13680) to C.J.B. We thank T. Flotte for encouragement and advice regarding genome editing strategies, and similarly appreciate the support of S. Jones and P. Newburger. We thank T. Collingwood for initial discussions regarding this project, and the George Daley laboratory (Harvard) for the Down's syndrome iPS cell line. L. Lizotte, Z. Matijasevic, K. Smith and E. Swanson provided various assistance. M. S. Kobor and L. Lam (Kobor laboratory) assisted with methylation analysis. D.M.C. is supported by an NIH fellowship 1F32CA154086 and B.R.C. (O. Rando laboratory) is supported by NIH training grant 2T32HD007439 (G. Witman, PI).

Author Contributions J.J., with the assistance of Y.J., designed and produced all constructs, edited all cell lines, and designed and performed most experiments. J.B.L., J.J. and L.L.H. were the main contributors to designing experiments and interpreting results. J.B.L., J.J., L.L.H. and F.D.U. wrote the manuscript. F.D.U., P.D.G. and G.J.C. engineered and validated ZFNs. J.R.P. performed Cel1 and Southern analysis. J.-C.C. performed SNP analysis, characterized three sub-clones, and helped with proliferation experiments. J.J. and Y.J., with help from J.-C.C., M.B., H.J.K. and L.L.H., carried out initial screening of targeted iPS cell sub-clones. H.J.K. edited and characterized primary Down's syndrome fibroblast line. A.M.C. and C.J.B. carried out DNA methylation analysis and provided *XIST* cDNA. J.J. and F.D.U. prepared the microarray library. D.M.C. and B.R.C. analysed microarray data with help from D.A.S., D.Y.G. and E.J.R.

Author Information Microarray data for 27 samples is deposited in GEO under accession number GSE47014. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B.L. (Jeanne.Lawrence@umassmed.edu) or F.D.U. (furnov@sangamo.com), requests for ZFNs.

METHODS

Cell culture. HT1080 TetR cells (Invitrogen) and female Down's syndrome human primary fibroblast line (Coriell) (AG13902) were cultured as recommended by the supplier. Down's syndrome iPS cell parental line (DS1-iPS4) was provided by G. Q. Daley (Children's Hospital Boston)²⁰ and maintained on irradiated mouse embryonic fibroblasts (iMEFs) (R&D Systems, PSC001) in hiPSC medium containing DMEM/F12 supplemented with 20% knockout serum replacement (Invitrogen), 1 mM glutamine (Invitrogen), 100 µM non-essential amino acids (Invitrogen), 100 µM β-mercaptoethanol (Sigma) and 10 ng ml⁻¹ FGF-β (Invitrogen, PHG0024). Cultures were passaged every 5–7 days with 1 mg ml⁻¹ of collagenase type IV (Invitrogen).

ZFN design. ZFNs against the human *AAVS1* locus on chromosome 19 have been previously described¹⁹. ZFNs against the *DYRK1A* locus were designed using an archive of pre-validated zinc finger modules^{18,29}, and validated for genome editing activity by transfection into K562 cells and Surveyor endonuclease-based measurement of endogenous locus disruption ('Cell' ^{30,31}) exactly as described²⁹. Southern blotting for targeted gene addition was performed exactly as described^{17,32} on Sphi-I-digested genomic DNA probed with a fragment corresponding to positions Chr21:38825803 + 38826056 (hg19).

XIST and rtTA/puro plasmid construction. Fourteen-kilobase human *XIST* cDNA, a splicing isoform of full-length *XIST* cDNA, was subcloned into pTRE3G (Clontech, catalogue no. 631167). Two homologous arms (left arm, 690 bp; right arm, 508 bp) of *DYRK1A* gene on chromosome 21 were amplified by PCR from primary Down's syndrome fibroblasts (AG13902) (Coriell) and cloned into the pTRE3G vector (human chromosome 21 *DYRK1A* left arm primers: forward 5'-GCCGTATACCATTAACCTCTTACTGTTTC-3', reverse 5'-TCTGTATACGTAA ACTGGCAAAGGGTGG-3'; human chromosome 21 *DYRK1A* right arm primers: forward 5'-ATTTCGGCGAACGGGTGATGAGCAGGCTGT-3', reverse 5'-CCGTCGCGAAAACCAGAAAGTATTCTCAG-3'). The pEF1α-3G rtTA-pA cassette from pEF1α-Tet3G vector (Clontech) was subcloned into a plasmid for targeted gene addition to the *PPPIR12C/AAVS1* locus¹⁹, which contains a unique HindIII site flanked by two 800-bp stretches of homology to the ZFN-specified position in the genome.

Dual-targeted addition of human Down's syndrome iPS cells and generation of stable targeted clones. The Down's syndrome iPS cell line was cultured in 10 µM of Rho-associated protein kinases (ROCK) inhibitor (Calbiochem; Y27632) 24 h before electroporation. Single cells (1×10^7) were collected using TryPLE select (Invitrogen), re-suspended in 1× PBS and electroporated with a total of 55 µg DNA including five plasmids (*XIST*, *DYRK1A* ZFN1, *DYRK1A* ZFN2, rtTA/puro and *AAVS1* ZFN) with both 3:1 and 5:1 ratios of *XIST*:rtTA/puro. The electroporation conditions were 220 V and 750 µF (BioRad Gene Pulser II System). Cells were subsequently plated on puromycin-resistant DR4 MEF feeders (Open Biosystems, catalogue no. MES3948) in hiPSC medium supplemented with ROCK inhibitor for the first 24 h. Over 300 colonies remained after 12 days of 0.4 µg ml⁻¹ puromycin selection and 245 randomly chosen individual colonies across 36 pooled wells were examined by interphase DNA/RNA FISH for the presence and expression of *XIST*, correct targeting and retention of trisomy (because some subclones lacked *XIST* or showed just two *DYRK1A* DNA signals). Over 100 individual clones were isolated and characterized, and those of interest, containing targeted *XIST* on one of three *DYRK1A* loci, were frozen. Six single target clones with good pluripotent morphology, OCT4 positive staining, correct targeting to one trisomic chromosome, and good *XIST* RNA paint were expanded for further characterization. One double and one triple target line, two non-target clones, and one disomic clone were also isolated and frozen. Targeting and correct chromosome number (47) was confirmed by interphase and metaphase FISH and genome integrity was confirmed by high-resolution G-band karyotype and CGH array.

Chromosome preparation. iPS cells were treated with 100 ng ml⁻¹ KaryoMAX colcemid (Invitrogen) for 2–4 h at 37 °C in a 5% CO₂ incubator. Cells were trypsinized, treated with hypotonic solution, and fixed with methanol:acetic acid (3:1). Metaphases were spread on microscope slides, and at least 20 analysed per clone. Karyotype analysis was done on pro-metaphase chromosomes using Standard Giemsa-trypsin G band methods.

CGH array. CGH was performed in the Cytogenetics Laboratory at University of Massachusetts Medical School. Genomic Microarray analysis using University of Massachusetts Genomic Microarray platform (Human Genome Build hg19) was performed with 1 µg of DNA. The array contains approximately 180,000 oligonucleotides (60-mers) that represent coding and non-coding human sequences and high-density coverage for clinically relevant deletion/duplication syndromes and the telomeric and pericentromeric regions of the genome. Data were analysed by BlueFuse Multi, v3.1 (BlueGnome, Ltd).

DNA/RNA FISH and immunostaining. DNA and RNA FISH were carried out as previously described^{10,15,16,33}. The *XIST* probe is a cloned 14-kb *XIST* cDNA

(the same sequence as *XIST* transgene in Fig. 1b) in pGEM-7Zf(+) (Promega). Six chromosome 21 gene probes are BACs from BACPAC Resources (*DYRK1A*, RP11-105O24; *APP*, RP11-910G8; *USP25*, RP11-840D8; *CXADR*, RP11-115O14; *ITSN1*, RP11-1033C16; *COL18A1*, RP11-867O18). DNA probes were labelled by nick translation with either biotin-11-dUTP or digoxigenin-16-dUTP (Roche). In simultaneous DNA/RNA FISH (interphase targeting assay), cellular DNA was denatured and hybridization performed without eliminating RNA and also treated with 2 U µl⁻¹ of RNasin Plus RNase inhibitor (Promega). For immunostaining with RNA FISH, cells were immunostained first with RNasin Plus and fixed in 4% paraformaldehyde before RNA FISH. Antibodies were as follows: H3K27me3 (Millipore, 07-449), UbH2A (Cell Signaling, 8240), H4K20me (Abcam, ab9051), macroH2A (Millipore, 07-219), OCT4 (Santa Cruz, sc-9081), PAX6 (Stemgent, 09-0075), SOX1 (R&D Systems, AF3369).

Allele-specific SNP analysis. Primers were designed to amplify 3' untranslated regions of chromosome 21 genes reported to contain SNPs (Supplementary Table 3). Total cDNA was used from three transgenic clones with and without *XIST* induction for 22 d. RT-PCR products were sequenced by GENEWIZ. Of ~10 genes examined, four were heterozygous and informative in the patient Down's syndrome iPS cell line used here.

Microarray analysis. Three independently targeted subclones plus the parental chromosome 21 trisomic (non-targeted) iPS cell line were grown with or without doxycycline (2 µg ml⁻¹) for 22 d. Normal male iPS cell and disomic isogenic lines were also cultured for 22 d and total RNA was extracted with a High Pure RNA extraction kit (Roche) in triplicate for each, processed with a Gene Chip 3' IVT express kit (Affymetrix), and hybridized to Affymetrix human gene expression PrimeView arrays. Array normalization was performed with Affymetrix Expression Console Software with Robust Multichip Analysis (RMA)³⁴. Probe sets with the top 60% of signal values were considered present and 'expressed' and were used for all further analysis. Data in Fig. 4 has no other threshold applied. When designated, a gene expression change significance threshold was applied using a two-tailed *t*-test comparing samples with or without doxycycline in triplicate ($n = 3$) (Supplementary Table 2, $P < 0.01$). For the ~200 genes found to significantly change in all three clones (in text), a *t*-test with $P < 0.001$ was applied.

Microarray data interpretation. Using extraction-based methods, changes on just one of three alleles (from the *XIST*-bearing chromosome) will be diluted by the other two. If all three chromosomes are fully expressed, this would predict a 33% reduction in chromosome 21 expression levels per cell when one chromosome 21 is fully silenced. However, 33% would apply only if chromosome 21 genes are fully overexpressed to start, and previous evidence and results in this study show this is not the case for many genes. Previous microarray studies have analysed expression levels of chromosome 21 in Down's syndrome patient cells, although such analyses are hampered by the extensive genetic and epigenetic differences between any two individuals³. The fraction of chromosome 21 genes detected as overexpressed varies with the study and tissue, but generally is in the 19–36% range^{3,22}, with individual gene increases often in the ~1.2–1.4 range (less than the theoretical 1.5). For example, one study of Down's syndrome embryoid bodies showed that only 6–15% of genes appeared significantly upregulated, but this was comparing non-isogenic samples of different ES cell isolates²².

Our trisomy correction system allows direct comparison of the same cells grown in identical parallel cultures, with and without *XIST*-mediated chromosome silencing. Our data show a ~20% reduction in chromosome 21 expression overall; importantly, this level of reduction is seen either when the third chromosome is silenced in trisomic cells, or when disomic and trisomic cells are compared. This 20% reduction represents an average per cell for all three chromosomes, but corresponds to a 60% reduction in expression for just one chromosome 21 (the one silenced by *XIST* RNA, as shown here).

Apart from our goal here of trisomy dosage compensation, these results add significantly to understanding the extent of chromosome 21 overexpression in Down's syndrome, by providing a more comprehensive analysis that shows that expression of most genes is increased, but less than the theoretical 1.5 fold.

qRT-PCR. qRT-PCR was performed for eight downregulated chromosome 21 genes determined by microarray on a Bio-Rad MyIQ real-time PCR detection system in triplicate for clone 3 with/without doxycycline treatment for 22 d. The β-actin gene was used as an internal standard for calculation of expression levels. Primers for eight chromosome 21 genes and β-actin were described in Supplementary Table 4.

DNA methylation analysis. The parental line and two independent targeted lines were grown with and without doxycycline for 22 d, in duplicate cultures. Genomic DNA was extracted using PureLink Genomic DNA mini kit (Invitrogen) and 750 ng bisulphite modified with the Alternative Incubation Conditions from the EZ DNA methylation kit (Zymo Research). 160 ng of bisulphite DNA was amplified, fragmented and hybridized to Illumina Infinium HumanMethylation450

array following the standard protocol as outlined in the user guide. CpG islands were defined as high and intermediate CpG densities using the CpG density classifications based on those used previously³⁵. The program CpGIE was used to locate HC and IC islands on the X chromosome and chromosomes 21 and 22. When multiple probes in CpG islands were associated with the same TSS, an average genic methylation value was calculated. These average genic values were compared before and after doxycycline induction using the Mann–Whitney *U*-test. Analysis was based on CpG islands within promoters of 143 chromosome 21 genes (Fig. 4c).

The average methylation value was 6% on chromosome 21 before *XIST* induction, and increased to 20–21% in both subclones after induction. Because any methylation increase on the transgenic chromosome would be diluted by the presence of three chromosome 21 copies, this suggests the range of 60% methylation on the one *XIST*-coated chromosome, which is within the range seen for the inactive X chromosome²⁴.

Cell proliferation analysis. Eight different iPS cell lines (parental line, one non-targeted subclone, and six independent targeted subclones) were passaged onto 6-well plates at equal cell densities per well of each line and grown with or without doxycycline for 7 d. At least four replicates of each line were analysed in two independent experiments. Rigorous measures were taken to minimize and control for any minor variations in seeding densities of iPS cells, which cannot be plated as single cell suspensions. First, the analysis was done twice for six different transgenic clones, in each case comparing triplicate plates of corrected versus not corrected (doxycycline versus no doxycycline). To avoid differences in plating efficiencies of doxycycline and no doxycycline cells, we performed the experiments over a time course that did not require passage. For each of the six transgenic clones, the parental line and one negative control (non-targeted) subclone, a single well of Down's syndrome iPS cells (without doxycycline) was used to generate a cell suspension (cells and small disaggregated clumps). Next, equal aliquots of the cell suspension were plated into each of six wells four times (not relying on one measurement but the average of four for seeding each well). After plating, doxycycline was added to three of the six wells, and the cultures were maintained for 7 d. For images, plates were fixed, stained with 1 mg ml⁻¹ crystal violet (Sigma) in 70% ethanol for 30 min and scanned to generate TIFF images. For cell counts, single cells were collected by TryPLE select and counted using Beckman Coulter Z1 Particle Counter.

Differentiation of neural progenitors and irreversibility in cortical neurons. For differentiation, independent *XIST*-transgenic iPS cell clones and the parental Down's syndrome iPS cell line were dissociated with Accutase (Innovative Cell Technologies) and 4×10^5 single cells were plated on Matrigel-coated 6-well plates in mTeSR1 medium (Stemcell technologies). Once the cell culture reached 90–100% confluence, neural induction was initiated by changing the culture medium to neural induction medium, a 1:1 mixture of N2- and B27-containing media supplemented with 500 ng ml⁻¹ noggin (R&D Systems), 10 µM SB431542 (Tocris Bioscience), and 1 µM retinoic acid (Sigma, catalogue no. R2625), with/without treatment of doxycycline for the specified times. The neural rosettes were counted and their diameter measured for at least 300 rosettes (sampled in random areas from triplicate dishes). At day 14, the doxycycline-induced culture had an average rosette diameter of 142 µm ± 0.55 µm in clone 1 and 141 µm ± 3.49 µm in clone 3. Rosettes could not be measured at the same time point in the uncorrected culture, as they had not formed. At day 17, the uncorrected culture had

neural rosettes of similar number and size for both clones 1 (140 µm ± 0.87 µm) and 3 (140 µm ± 1.09 µm). The corrected culture could not be accurately compared for day 17 because the rosettes had become so mature and often had merged. After 17 d, neural rosettes were collected by dissociation with dispase and replated on poly-ornithine and laminin-coated plastic dishes in N2- and B27-containing media including 20 ng ml⁻¹ FGF2. After a further 2 d, FGF2 was withdrawn to promote differentiation of cortical neurons. To test for the irreversibility of silencing, two independent clones were differentiated to cortical neurons in the presence of doxycycline for 70 days to initiate silencing. They were then split into parallel cultures grown with and without doxycycline for another 30 days, and *XIST* and APP expression analysed by RNA FISH.

Targeted addition to primary fibroblasts. We used non-immortalized primary human female Down's syndrome fibroblasts, which like all primary fibroblasts have a limited lifespan in culture (potentially more limited for Down's syndrome fibroblasts). We reasoned that the robustness of ZFN-driven editing, combined with reduction to disomy for the *DYRK1A* gene, may make it possible to observe some edited cells before they senesce. We used a transgene carrying near full-length (~14 kb) *XIST* cDNA under a TetO₂ inducible promoter, and a selectable marker on the same construct, with ~600-bp homology arms to the *DYRK1A* gene (vector is ~21 kb, with a ~17-kb insert) (data not shown). When introduced without the Tet-repressor construct, the TetO₂ CMV promoter is constitutively active. Two ZFN-containing vectors and the 21-kb *XIST* transgene were transfected into primary DS fibroblasts (Coriell AG13902) using Stemfect polymer (Stemgent) (10:1 ratio of *XIST* to ZFN, and 13 µg DNA to 1.3 µl Stemfect per well of 6-well plate). The frequency of stable integrants was such that a sparse monolayer of transgenic fibroblasts emerged, rather than a few individual colonies following selection with hygromycin (75 µg ml⁻¹). The pooled population of selected cells was analysed by FISH and immunostaining for targeting, *XIST* expression and heterochromatin marks. *XIST* RNA was observed over the *DYRK1A* locus in ~74% of cells, indicating accurate transgene targeting, which was also verified by metaphase FISH (Supplementary Fig. 9c). In many cells there was notable enrichment of H3K27me, UbH2A and H4K20me heterochromatic marks (Supplementary Fig. 9d). Owing to the limited lifespan of primary cells and the progressive silencing of the CMV promoter used in this construct, these cells were not more fully characterized.

29. Doyon, J. B. et al. Rapid and efficient clathrin-mediated endocytosis revealed in genome-edited mammalian cells. *Nature Cell Biol.* **13**, 331–337 (2011).
30. Miller, J. C. et al. An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnol.* **25**, 778–785 (2007).
31. Guschin, D. Y. et al. A rapid and general assay for monitoring endogenous gene modification. *Methods Mol. Biol.* **649**, 247–256 (2010).
32. Urnov, F. D. et al. Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature* **435**, 646–651 (2005).
33. Byron, M., Hall, L. L. & Lawrence, J. B. A multifaceted FISH approach to study endogenous RNAs and DNAs in native nuclear and cell structures. *Curr. Protoc. Hum. Gen.* Chapter 4, Unit 4.15 (2013).
34. Irizarry, R. A. et al. Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
35. Weber, M. et al. Distribution, silencing potential and evolutionary impact of promoter DNA methylation in the human genome. *Nature Genet.* **39**, 457–466 (2007).

Coordination of bacterial proteome with metabolism by cyclic AMP signalling

Conghui You^{1,2}, Hiroyuki Okano^{1,2*}, Sheng Hui^{1,3*}, Zhongge Zhang², Minsu Kim¹, Carl W. Gunderson², Yi-Ping Wang⁴, Peter Lenz⁵, Dalai Yan⁶ & Terence Hwa^{1,2,3}

The cyclic AMP (cAMP)-dependent catabolite repression effect in *Escherichia coli* is among the most intensely studied regulatory processes in biology. However, the physiological function(s) of cAMP signalling and its molecular triggers remain elusive. Here we use a quantitative physiological approach to show that cAMP signalling tightly coordinates the expression of catabolic proteins with biosynthetic and ribosomal proteins, in accordance with the cellular metabolic needs during exponential growth. The expression of carbon catabolic genes increased linearly with decreasing growth rates upon limitation of carbon influx, but decreased linearly with decreasing growth rate upon limitation of nitrogen or sulphur influx. In contrast, the expression of biosynthetic genes showed the opposite linear growth-rate dependence as the catabolic genes. A coarse-grained mathematical model provides a quantitative framework for understanding and predicting gene expression responses to catabolic and anabolic limitations. A scheme of integral feedback control featuring the inhibition of cAMP signalling by metabolic precursors is proposed and validated. These results reveal a key physiological role of cAMP-dependent catabolite repression: to ensure that proteomic resources are spent on distinct metabolic sectors as needed in different nutrient environments. Our findings underscore the power of quantitative physiology in unravelling the underlying functions of complex molecular signalling networks.

Biological organisms use a myriad of signalling pathways to monitor the environment and adjust their genetic programs in accordance with environmental changes^{1–4}. Understanding how the signalling systems perceive the environment and orchestrate the downstream genetic changes is a grand challenge of systems biology^{5–8}. One of the earliest signalling systems ever discovered in modern biology is the cyclic AMP (cAMP)-dependent pathway in *E. coli*^{9,10}. This system is known to mediate carbon catabolite repression (CCR)¹¹, a ubiquitous phenomenon among microorganisms whereby the synthesis of catabolic proteins is inhibited when growing on glucose or other rapidly metabolizable sugars^{12,13}. In *E. coli*, it was long established that the uptake of glucose inhibited the synthesis of cAMP¹⁰, which is required for the expression of many catabolic genes through its activation of the pleiotropic regulator Crp (the cAMP receptor protein)¹⁴.

Despite extensive studies on this cAMP signalling pathway, a number of basic issues remain unanswered even for this well-known system: although the inhibitory effect of glucose uptake on cAMP synthesis via the phosphotransferase system (PTS) is well established^{15,16} (Supplementary Fig. 1), the growth of *E. coli* cells on various PTS-independent sugars also showed reduced cAMP levels^{17–19}. Moreover, limitation of nitrogen, phosphorous, and other elements also led to much reduced cAMP levels^{20–22}. Elucidating the physiological signal(s) that triggered cAMP signalling in those conditions was in fact the pursuit of Monod's last paper²³, but these signal(s) remain elusive. Also, it is unclear to what extent cAMP signalling is intended for implementing CCR²⁴, as hierarchical carbon usage²⁵, a behaviour widely associated with CCR¹⁶, was shown to be independent of cAMP in several studies^{26,27}. Thus, the true physiological function of cAMP signalling in *E. coli* remains open nearly 50 years after its discovery. In this study, we describe a top-down approach which first addresses quantitatively the physiological function

of cAMP signalling—not for CCR *per se* as we will show, but for the coordination of metabolism with proteomic resource allocation. This knowledge is then used as a guide to reveal the signalling strategy and mechanism by which *E. coli* cells trigger cAMP signalling.

Catabolic genes show linear response

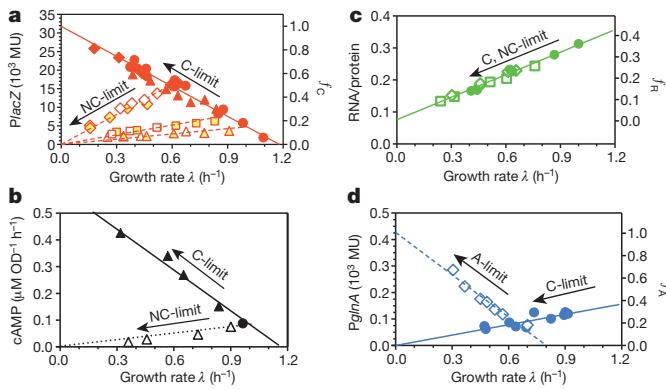
To characterize the physiological consequences of cAMP-dependent signalling, we used the well-studied *lac* system of *E. coli*, one of the many catabolic operons activated by the Crp-cAMP complex^{14,28}. Wild-type *E. coli* K-12 cells were grown in minimal medium batch culture with saturating amount of one of many carbon sources, with the Lac repressor (LacI) deactivated by the inducer isopropyl-β-D-thiogalactoside (IPTG). As such, the native LacZ expression indicated the degree of cAMP signalling²⁹. Supplementary Table 1 lists the LacZ expression level (L) together with the growth rate (λ) for cells growing in each medium. A scatter plot of the data shows a striking linear relation (solid circles in Fig. 1a). The same relation is obtained by limiting the carbon influx, whether by titrating the lactose permease for cells growing on lactose (Supplementary Table 2, solid triangles in Fig. 1a), or by titrating the glycerol uptake system for cells growing on glycerol (Supplementary Table 3, solid diamonds in Fig. 1a); see Supplementary Figs 2, 3 and Supplementary Table 4 for strain details. The red solid line is the best-fit of all data to the form

$$L = L_{\max}(1 - \lambda/\lambda_C) \quad (1)$$

referred to below as the 'C-line'; see Supplementary Table 5 for fitted values of the x -intercept λ_C and y -intercept L_{\max} . Similar linear relations are seen (Supplementary Fig. 4) for *lac* in various other strains of *E. coli*

¹Department of Physics, University of California at San Diego, La Jolla, California 92093-0374, USA. ²Section of Molecular Biology, Division of Biological Sciences, University of California at San Diego, La Jolla, California 92093 USA. ³Center for Theoretical Biological Physics, University of California at San Diego, La Jolla, California 92093-0374 USA. ⁴State Key Laboratory of Protein and Plant Gene Research, College of Life Sciences, Peking University, Beijing 100871, China. ⁵Department of Physics and Center for Synthetic Microbiology, University of Marburg, 35032 Marburg, Germany. ⁶Department of Microbiology and Immunology, Indiana University School of Medicine, Indianapolis, Indiana 46202, USA.

*These authors contributed equally to this work.



	Strain	Description	Carbon	LacZ	cAMP	RNA/protein
C-limitation	NCM3722	WT	Various	Supplementary Table 1 (5)	● Supplementary Table 8	● Supplementary Table 10 (14)
NQ381	Titratable LacY	Lactose	Supplementary Table 2 (5)	▲ Supplementary Table 8		
NQ399	Titratable GlpFK	Glycerol	Supplementary Table 3 (5)	◆ Supplementary Table 8		
NQ158	PglmA-lacZ	Various	Supplementary Table 12 (15)	● Supplementary Table 9		
NQ34	Titratable GDH	Glucose	Supplementary Table 7 (5)	□ Supplementary Table 9	□ Supplementary Table 11 (14)	□ Supplementary Table 11 (14)
		Glycerol	Supplementary Table 7 (5)	◇ Supplementary Table 9	◇ Supplementary Table 11 (14)	
NC(A)-limitation	NCM3722	N-chemostat	Lactose	Supplementary Table 6 (5)	● Supplementary Table 8	
NQ354	N-chemostat	Glucose	Supplementary Table 6 (5)	□ Supplementary Table 8		
NQ354	S-chemostat	Glycerol	Supplementary Table 6 (5)	◇ Supplementary Table 8		
NQ477	PglmA-lacZ	Glycerol	Supplementary Table 13 (15)	◊ Supplementary Table 8		

Figure 1 | Catabolic and biosynthetic gene expression under nutrient limitations. For clarity, growth conditions and depository of data are summarized in the legend table. Number in parentheses indicate the Supplementary Tables containing the parameters of the best-fit lines.

a, Correlation of *PlacZ* expression with the growth rate under C-limitations (solid symbols) and NC-limitations (open symbols). The left y-axis shows *PlacZ* expression per OD₆₀₀ (Miller unit or MU), which is proportional to per total protein; see Supplementary Methods and Supplementary Fig. 14. The right y-axis shows the equivalence in normalized fractions f_C (Supplementary Note 3). **b**, Internal cAMP level as indicated by the cAMP excretion rate (Supplementary Fig. 11) for cells grown under C- and NC-limitations (filled and open symbols, respectively). **c**, RNA/protein ratio for both C- and NC-limited growth (filled and open symbols, respectively). The right y-axis shows the equivalence in normalized fractions f_R (Supplementary Note 3). **d**, *PglmA-lacZ* expression under C- and A-limited growth (filled and open symbols, respectively). The right y-axis shows the equivalence in normalized fractions f_A (Supplementary Note 3).

characterized here and previously^{20,30,31}, and for different catabolic promoters (solid symbols in Supplementary Fig. 5).

The data above indicate that the C-line is a common response pattern shown by the catabolic genes—a gradual increase upon reduction in carbon influx or for less rapidly metabolized carbon compounds. We refer to such carbon-dependent mode of growth limitation as ‘C-limitation’. To show that the C-line is a response specific to changes in carbon catabolism rather than to general growth parameters^{32,33}, we modulated cell growth by various non-carbon (NC) modes of growth limitation, including continuous culture with ammonium or sulphate limitation (Supplementary Table 6), and batch culture using a titratable nitrogen uptake system (Supplementary Table 7). The latter system bypasses the highly regulated nitrogen response system³⁴, creating a situation in which the formation of amino acids from their respective carbon can be directly titrated by the expression of glutamate dehydrogenase (GDH); see Supplementary Fig. 6. The expression of the *lacZ* promoter (*PlacZ*) obtained for these modes of NC-limited growth is plotted against the growth rates in Fig. 1a and Supplementary Fig. 7

for several exemplar carbon sources (open symbols). In each case, a positive linear relationship was observed (dashed lines). Similar relations are seen for several other catabolic promoters (open symbols in Supplementary Fig. 5d, e). We refer to these positive linear relations as the ‘NC-lines’. Both C- and NC-lines also described the expression of a *PlacZ-gfp* fusion reporter in wild-type cells grown in microfluidic devices with rapid medium exchange³⁵ to maintain low nutrient levels and remove excretion products (Supplementary Fig. 8).

We next examined the dependence of the C- and NC-lines on Crp-cAMP-mediated activation. Upon replacing the *lacZ* promoter by the *lacUV5* promoter, whose independence of Crp-cAMP is well established³⁶, LacZ expression in these strains showed little growth-rate dependence under both C- and NC-limitations (Supplementary Fig. 9). Additionally, we characterized the expression of a synthetic *lacZ* promoter with the Crp site scrambled, and found only a weak growth-rate dependence, resembling that previously reported for constitutive gene expression³³, under both C- and NC-limitations (Supplementary Fig. 10). Thus both the C- and NC-lines rely completely on Crp-cAMP-mediated gene regulation.

We also characterized the cAMP excretion rate (Supplementary Fig. 11), taken to reflect the internal cAMP concentration¹⁷. It is seen to correlate well with *PlacZ* expression in the respective medium for both C- and NC-limited growth (Supplementary Fig. 12, Supplementary Tables 8 and 9). Moreover, scatter plot of cAMP excretion rates against the corresponding growth rates (Fig. 1b) resembles the C- and NC-lines of *PlacZ* (Fig. 1a). Together, these data strongly indicate that the C- and NC-lines are two sides of the same coin, both regulated by Crp-cAMP primarily through cAMP signalling.

Opposing linear response by anabolic genes

To understand the linearity of catabolic gene expression with the growth rate, we recall a linear relation between the cellular ribosome content and the growth rate for exponentially growing bacteria^{33,37,38}. This relation (a bacterial growth law), which arises due to the allocation of ribosomes as a key growth-limiting resource^{33,38}, is verified for our strains by characterizing the RNA/protein ratio (r) under both C- and NC-limited growth (Supplementary Tables 10 and 11, Fig. 1c). The striking contrast between the linear increase in catabolic gene expression (solid red symbols in Fig. 1a) and the concomitant linear decrease in the ribosome content (Fig. 1c) upon C-limitation (solid green symbols in Fig. 1c) indicates that these phenomena might share a common origin.

This line of qualitative reasoning predicts that, upon NC-limitation where the expression levels of catabolic and ribosomal genes both decrease linearly as fractions of total protein (open symbols in Fig. 1a, c), the expression of other genes (for example, biosynthetic genes) should increase linearly with the growth rate. We tested this prediction by characterizing the expression of *glnA*, which encodes the major ammonia assimilating protein glutamine synthetase³⁴ using the titratable GDH system under a fixed carbon source (glycerol). As expected, the expression level of the chromosomal *PglmA-lacZ* fusion reporter (open blue diamonds in Fig. 1d, denoted as G) had a negative linear correlation with growth rate; similar behaviours were seen for the expression of several other genes in amino acid synthesis pathways (Supplementary Fig. 13). These responses suggest a general amino acid limitation imposed by the titratable GDH system (Supplementary Fig. 6), referred to henceforth as ‘A-limitation’ and the resulting response as the ‘A-line’ (dashed blue line in Fig. 1d). The latter is described mathematically by

$$G = G_{\max}(1 - \lambda/\lambda_A) \quad (2)$$

with an x -intercept λ_A and a y -intercept G_{\max} (Supplementary Table 15). Interestingly, *PglmA-lacZ* also had a positive linear correlation with the growth rate in C-limited growth (solid blue line in Fig. 1d). Thus *E. coli*

perceives carbon limitation as a nitrogen excess, a feature reported qualitatively long ago in *Klebsiella*³⁹.

Quantifying the proteome fractions

To see whether the response of the catabolic genes can be quantitatively accounted for by changes in the biosynthetic genes and the ribosome content, we apply and extend a mathematical model of proteome partition introduced recently^{33,40} to relate various growth laws to the global regulation of gene expression. As illustrated in Fig. 2a and explained in detail in Supplementary Note 1, we propose that the proteome is comprised of several major ‘sectors’, that is, groups of genes whose expression share similar growth-rate dependences upon various modes of growth limitation. These include a catabolic sector (C) of mass fraction ϕ_C that increases specifically in response to C-limitation as exemplified by *lacZ* (Fig. 1a and Supplementary Fig. 5), an anabolic sector (A) of mass fraction ϕ_A that increases specifically in response to A-limitation as exemplified by *glnA* (Fig. 1d, Supplementary Fig. 13), and a ribosomal sector (R) of mass fraction ϕ_R that increases specifically in response to translational inhibition (Supplementary Fig. 15). To account for additional sectors which have not been specifically targeted by the various modes of growth limitation applied in this study, we propose an uninduced sector (U) of mass fraction ϕ_U which generally decreases with decreasing growth rate. Examples of the latter might include enzymes for sulphur assimilation and nucleotide synthesis (Supplementary Fig. 16).

As shown by the various reporters, we expect each of the above proteome sector j to contain a growth-rate dependent part $\Delta\phi_j(\lambda)$ and a growth-rate independent offset $\phi_{j,0}$, with $\phi_j(\lambda) = \phi_{j,0} + \Delta\phi_j(\lambda)$; see Supplementary Fig. 17. Additionally, there may exist a growth-rate independent sector (I) of mass fraction ϕ_I . Together, the growth-rate independent components can be lumped into an effective sector (Q) of mass fraction $\phi_Q = \phi_I + \sum_j \phi_{j,0}$ as suggested from a previous study³³, providing a cap $\phi_{\max} = 1 - \phi_Q$ on how large the specific responses of each sector can increase to, that is, $\sum_j \Delta\phi_j(\lambda) = \phi_{\max}$. We estimated $\phi_{\max} \approx 43\%$ using protein overexpression data (Supplementary Fig. 18 and Supplementary Note 2).

The constraint that the $\Delta\phi_j$ s add up to a constant (ϕ_{\max}) imposes quantitative relations between the responses of the various sectors upon different modes of growth limitation. This constraint is most simply expressed (Supplementary Note 3) in terms of a normalized fraction, $f_j(\lambda) = \Delta\phi_j(\lambda)/\phi_{\max}$ (see equation S21 in Supplementary Information) for each sector j , as

$$f_C(\lambda) + f_A(\lambda) + f_R(\lambda) + f_U(\lambda) = 1 \quad (3)$$

Assuming that changes in the expression levels of *PlacZ* and *PglmA* report changes in the C- and A-sectors, respectively, that is, $\Delta L \propto \Delta\phi_C(\lambda)$ and $\Delta G \propto \Delta\phi_A(\lambda)$, the normalized proteome sectors $f_C(\lambda)$ and $f_A(\lambda)$ can be estimated from the reporter expression levels using our model (Supplementary Fig. 17). The results are indicated on the right vertical scales in Fig. 1a, d. Similarly, $f_R(\lambda)$ can be obtained from the RNA/protein ratio $r(\lambda)$ (right vertical scale of Fig. 1c and Supplementary Fig. 15a). The values of these normalized fractions under C-limited growth are plotted together as coloured circles in Fig. 2b, and their best linear fits are shown as lines of corresponding colours. Since $f_C(\lambda) + f_A(\lambda) + f_R(\lambda) < 1$ (black open circles and line), we deduce from equation (3) the existence of an uninduced sector U whose abundance, $f_U(\lambda)$, is shown by the purple line, it is approximately 30% of the R-sector (green line).

Tight coordination of proteome sectors

The constraint equation (3) suggests that the growth-rate dependences of the C- and A-sectors are simply related, given that $f_R(\lambda)$ follows the growth law (Fig. 1c) and $f_U \approx 0.3f_R$. Applying this to A-limited growth, we can use the observed expression of *PlacZ* under

A-limitation (Fig. 1a) to predict *PglmA* expression under A-limitation; the result is in good agreement with the best fit as shown in Fig. 2c (black line), with glycerol as the carbon source. This is repeated for A-limitation on glucose, where we expect the smaller C-sector for glucose to lead to an expanded A-sector, and hence an increased intercept λ_A for *PglmA*. This prediction is verified in Supplementary Fig. 19, with the data quantitatively accounted for by the predicted A-line (black line). Similarly, the value of the intercept λ_C for *PlacZ* under C-limitation with a fixed degree of A-limitation is reduced to the expected range (Supplementary Fig. 20). See Supplementary Fig. 21 and Supplementary Table 16 for the direct comparison of the predicted and best-fit values of λ_C , λ_A . As one additional test of this picture, we applied sub-lethal amounts of antibiotics to probe the effect of translational limitation on the various proteome sectors (Supplementary Fig. 22). Whereas the R-sector increased linearly with decreasing growth rate³³, *PlacZ* and *PglmA* expressions both decreased linearly with decreasing growth rate as predicted, again in a manner that conforms quantitatively to the constraint equation (3).

These data sets (Fig. 2b, c, Supplementary Figs 19, 20) can also be displayed simultaneously on 3D plot (Fig. 2d); the data are seen to lie close to a plane, defined by equation (3) with $f_U \approx 0.3f_R$. This is an example of the Pareto surface⁴¹, which results generally from trade-off between competing objectives that cannot be simultaneously optimized. The linear dependences of these fractions on the growth rate further indicate that they arise from protein synthesis being the common

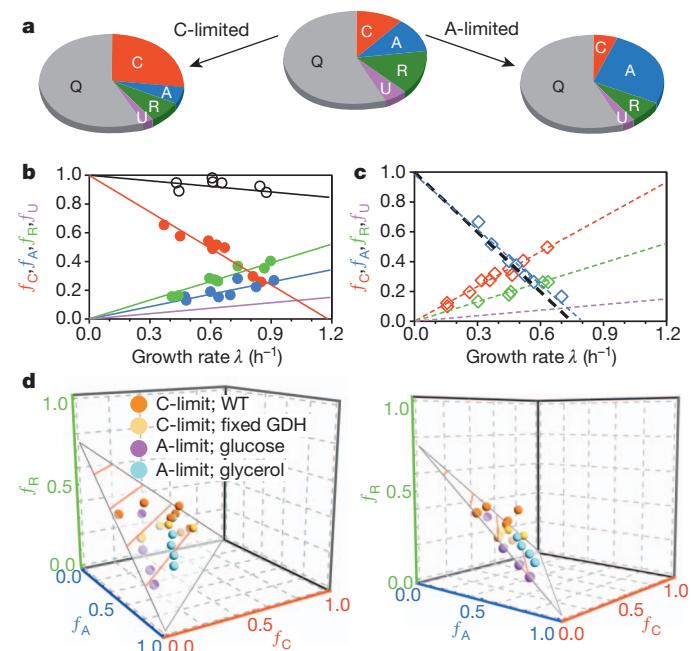


Figure 2 | Proteome fractions and the partition model. **a**, Illustration of the proteome partition model. Upon C-limitation, the C-sector increases and the A-, R-, U-sectors decrease, whereas upon A-limitation, the A-sector increases and the C-, R-, U-sectors decrease; see Supplementary Note 1. **b**, Normalized responses (f_C , f_A , f_R) indicated by the C-, A-, R- sector reporters upon C-limitation (solid red, green, blue circles respectively from Fig. 1a, c, d). The black circles show the sum $f_C + f_A + f_R$ at each growth rate; they decrease linearly with the growth rate (black line). The purple line is the predicted U-sector fraction f_U based on equation (3). **c**, f_C , f_A , f_R upon A-limitation with glycerol as the carbon source (open red, green, blue diamonds, respectively, from Fig. 1a, c, d). Taking f_U (purple line) to be the same as that in **b** ($f_U = 0.3f_R$), equation (3) predicts f_A to follow the black line, and blue line shows the best fit. **d**, f_C , f_A , f_R for the four sets of C- and A-limited growth conditions characterized (**b**, **c** and Supplementary Figs 19, 20) are plotted in a 3D plot. (Unavailable data for f_R , f_A are generated from the straight line fit.) Two views are shown. The data are seen to fall on the predicted surface (equation (3) with $f_U = 0.3f_R$).

bottleneck of cell growth. For example, catabolic proteins are needed to increase carbon influx while ribosomal and anabolic proteins are needed to increase biosynthesis, all of which depend on translation by ribosomes (Supplementary Note 4). This logic leads to an important physiological rationale for the pleiotropic regulation of catabolic genes by Crp-cAMP: given rapidly metabolizable carbons, cells reduce catabolic gene expression to increase the expression of ribosomal and anabolic genes for rapid cell growth.

Strategy of metabolic–proteome coordination

The opposing behaviours shown by the C- and A-sectors upon different modes of growth limitation (Figs 1 and 2) point strongly to the existence of control system(s). Figure 3a describes a simple regulatory scheme wherein a single signal can sense and eliminate imbalance between metabolic activities on the carbon and nitrogen side. In this coarse-grained scheme, the external carbon source is converted to the pool of carbon precursors (K), for example, α -ketoacids, which subsequently combine with ammonium to form amino acids (a) via transamination reactions (Supplementary Fig. 6). If the carbon influx (J_C) exceeds the nitrogen influx (J_N), then precursors (K) will accumulate^{42–44}. A signal from K to inhibit Crp-cAMP activity (dotted red line in Fig. 3a) can account for the observed responses by the catabolic genes (Fig. 1a and Supplementary Fig. 5), both for increases in J_C (relief of C-limitation) and decreases in J_N (A-limitation). An opposite signal from K could underlie changes in the A-sector (dashed blue arrow). Detailed mathematical analysis (Supplementary Note 5) shows how

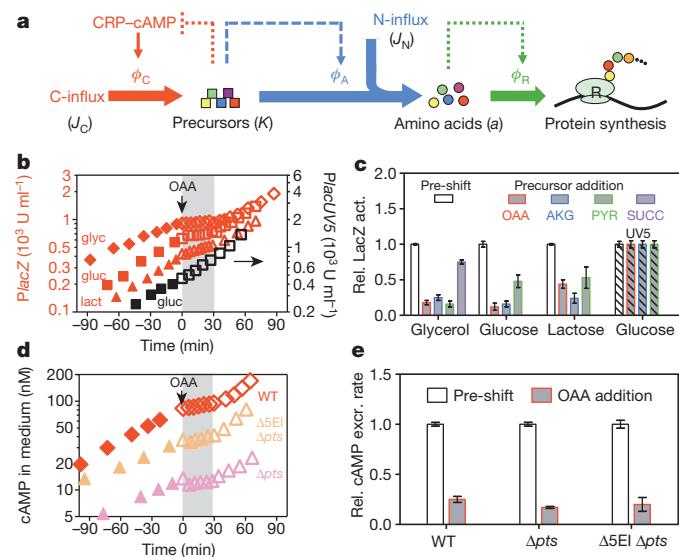


Figure 3 | Transient repression by metabolic precursors. **a**, A coarse-grained view of metabolism, focusing on the biosynthesis of amino acids from the carbon and nitrogen influxes (J_C and J_N respectively). Carbon precursors such as α -ketoacids (K) sense the difference between J_C and J_N . An integral feedback scheme using the regulation of catabolic and anabolic enzymes (ϕ_C and ϕ_N , respectively) by K can coordinate these metabolic sectors in a parameter-free manner; see Supplementary Note 5 for details. **b**, *PlacZ-lacZ* expression was characterized for wild-type NCM3722 cells grown exponentially in various carbon sources and with 1 mM IPTG to deactivate LacI. At time zero, 20 mM OAA was added; a transient repression period of ~ 30 min is shaded in grey. *PlacUV5-lacZ* expression in strain NQ1053 was characterized in the same way (black squares; right y-axis). **c**, LacZ expression levels before and during the repression period (Fig. 3b and Supplementary Figs 23–27) are summarized by the open and grey bars. Striped bars show the results of *PlacUV5-lacZ*. **d**, cAMP concentrations in the medium were monitored for wild-type cells grown in glycerol and the two PTS-deletion strains, NQ721 (Δpts) and NQ506 ($\Delta 5EI \Delta pts$), grown in lactose. 20 mM OAA was added at time zero. **e**, Relative cAMP excretion rates were quantified before and during the repression period; see Supplementary Method. In **c** and **e**, data were expressed as mean \pm s.e.m. ($n \geq 3$).

such a feedback loop (known as “integral feedback”^{35,45} since K is the time-integral of the flux difference between J_C and J_N being regulated), together with global constraints provided by proteome partition, can generate the C- and A-lines generically, independent of the detailed forms of the regulatory functions.

The proposed integral feedback strategy can be tested by adding a precursor to an exponentially growing culture. If the uptake of precursor significantly exceeds its metabolic turnover, then the transient build-up of the precursor should reduce the internal cAMP level and the expression of catabolic genes, until a later time when metabolic fluxes are re-equilibrated. We first tested oxaloacetate (OAA), a major metabolic precursor (an α -ketoacid) which is co-used with other carbons by wild-type cells to give faster steady-state growth (Supplementary Table 17). The addition of OAA to cells growing exponentially in glycerol, glucose or lactose strongly inhibited *PlacZ* expression (with LacI deactivated) for up to 30 min (Fig. 3b, c and Supplementary Fig. 23). This transient effect was Crp-cAMP-dependent because no effect was observed for *PlacUV5* (Fig. 3b, c) and another Crp-independent promoter (Supplementary Fig. 23f). The inhibitory effect is most probably exerted on cAMP synthesis, because the excretion of cAMP, which reflects the internal cAMP level¹⁷, was also inhibited over the same period upon OAA addition (Fig. 3d, e). Thus, OAA is a candidate precursor implementing the feedback inhibition depicted in the dotted red line of Fig. 3a.

Strong to moderate repressive effects were also observed following the addition of α -ketoglutarate (AKG) (Fig. 3c and Supplementary Figs 24, 25) and pyruvate (PYR) (Fig. 3c and Supplementary Fig. 26), both of which are α -ketoacids. In contrast, the addition of succinate (SUCC), which is a dicarboxylic acid like OAA and AKG, but is not an α -ketoacid, leads to only minor repression (Fig. 3c and Supplementary Fig. 27). Although our study does not rule out other possible causes of transient repression⁴⁶, the inhibitory effects produced by the α -ketoacids are much stronger than those by glucose and the non-hydrolysable compound methylthio- β -galactoside (TMG) investigated previously⁴⁷; see Supplementary Fig. 28. Finally, α -ketoacid addition also caused transient increase in the expression of the anabolic gene *glnA* (Supplementary Fig. 29a, b). This regulatory effect is independent of cAMP (Supplementary Fig. 29c, d, e), and is captured by the dashed blue arrow in Fig. 3a.

Mechanism of metabolic–proteome coordination

The cAMP signalling pathway linking metabolism to catabolic gene expression is commonly thought to comprise of the PTS proteins^{15,16} (Supplementary Fig. 1). Recently, AKG was reported to interact directly with enzyme I (EI) of the PTS proteins⁴³, making PTS proteins likely candidates for implementing the proposed feedback by precursors (dotted red line in Fig. 3a). However, strong transient repression was still observed upon the addition of α -ketoacids in strains with deletion of various PTS proteins, including one with the 3 major PTS proteins deleted (Δpts) and another with additionally all 5 EI proteins deleted ($\Delta 5EI \Delta pts$); see Fig. 4a, b and Supplementary Figs 30, 31. This transient effect was still accompanied by similar fold of reduction in cAMP excretion *in vivo* (Fig. 3d, e), and the PTS mutants in steady state still showed C-lines for C-limited growth (Fig. 4c). The C-lines of the PTS mutants (dashed lines) showed similar x -intercept (λ_C) as the wild type (red line), indicating that although PTS proteins do affect the degree of Crp-cAMP-mediated regulation of catabolic gene expression as is long known¹⁵, they are not necessary for the existence of this response (that is, the occurrence of the C-line).

Next we investigated directly the inhibitory effect of various candidate metabolites on cAMP synthesis, by *in vitro* assays with permeabilized cells (ref. 48 and Supplementary Methods). The rate of cAMP synthesis by adenylate cyclase was significantly reduced by several α -ketoacids but not by SUCC, in either PTS⁺ or PTS⁻ background (Fig. 4d and Supplementary Fig. 32). Thus, the simplest regulatory route consistent with all data are to have precursors such as

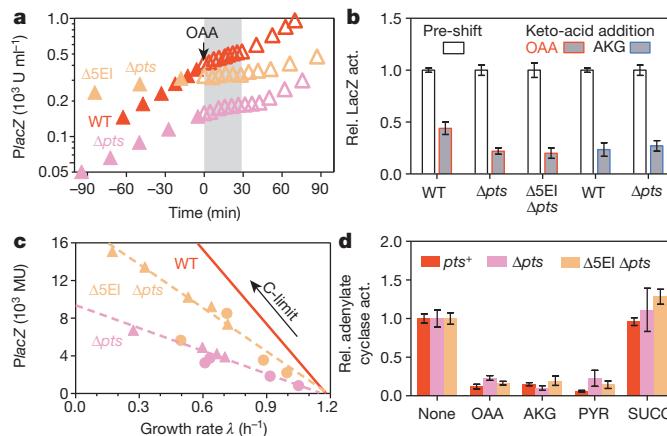


Figure 4 | Mechanism of cAMP-dependent signalling. **a**, OAA transiently repressed *PlacZ-lacZ* expression in PTS-deleted cells, NQ721 (Δpts) and NQ506 ($\Delta\text{5EI}\Delta\text{pts}$), grown exponentially in lactose. 20 mM OAA was supplied at time zero. The response of WT cells to OAA in Fig. 3b (red triangles) was plotted for comparison. **b**, The magnitudes of transient repression by OAA and AKG were quantified as in Fig. 3c, based on the data in Fig. 4a and Supplementary Figs 30, 31. **c**, Steady state *PlacZ-lacZ* expression in PTS mutants under various modes of C-limitations; see Supplementary Table 18 for data and conditions. Dashed lines show the best linear fit (Supplementary Table 19). The C-line of Fig. 1a is shown in red for reference. **d**, *in vitro* adenylate cyclase activities in strains NQ385 (pts^+), NQ976 (Δpts) and NQ977 ($\Delta\text{5EI}\Delta\text{pts}$) were assayed with or without 10 mM of various candidate inhibitors; see Supplementary Methods for details. These strains are also deleted of the cAMP phosphodiesterase which is not primary to signalling in cAMP-dependent CCR; see Supplementary Fig. 33. In **b** and **d**, data were expressed as mean \pm s.e.m. ($n \geq 3$).

OAA inhibiting cAMP synthesis by adenylate cyclase (Supplementary Fig. 34), either directly or through the help of a mediator protein. Simple inhibition kinetics was found for OAA, with half-inhibition concentrations in the sub-millimolar range (Supplementary Fig. 35).

Discussion

cAMP-mediated CCR is one of the most intensely studied regulatory systems in biology. Results described here indicate that the physiological function of cAMP signalling is not centred on carbon metabolism per se as commonly thought; rather this pathway coordinates the allocation of proteomic resources (Fig. 2a) with different metabolic demands in different nutrient environments. This global regulatory process is directed by an integral feedback system (Fig. 3a), driven by cAMP signals modulated by metabolic precursors such as several α -ketoacids. These precursors can be identified as the ‘catabolites’, the accumulation of which was proposed to induce CCR over 50 years ago¹¹.

The complexity of molecular interactions underlying this signalling system made it difficult to elucidate its physiological function in a bottom-up approach despite the wealth of molecular data known for this classic system. However, the top-down physiological approach described here was able to reveal simple quantitative relations between gene expression and growth rates; these relations provided key insight on the physiological function of cAMP signalling, leading to the elucidation of molecular strategy and implementation. The effectiveness of this approach provides a conceptual framework to pursue the use of quantitative physiological approaches to elucidate other complex signalling networks.

METHODS SUMMARY

Material and method including strain construction, cell culture, LacZ activity assay, cAMP excretion rate assay, RNA/protein ratio assay, *in vitro* adenylate cyclase activity assay and theoretical analysis are provided in detail in Supplementary Information.

Received 10 February 2012; accepted 11 July 2013.

Published online 7 August 2013.

- Laub, M. T. & Goulian, M. Specificity in two-component signal transduction pathways. *Annu. Rev. Genet.* **41**, 121–145 (2007).
- Potrykus, K. & Cashel, M. (p)ppGpp: still magical? *Annu. Rev. Microbiol.* **62**, 35–51 (2008).
- Hengge, R. Principles of c-di-GMP signalling in bacteria. *Nature Rev. Microbiol.* **7**, 263–273 (2009).
- Porter, S. L., Wadhams, G. H. & Armitage, J. P. Signal processing in complex chemotaxis pathways. *Nature Rev. Microbiol.* **9**, 153–165 (2011).
- Brent, R. Cell signalling: what is the signal and what information does it carry? *FEBS Lett.* **583**, 4019–4024 (2009).
- Purvis, J. E. *et al.* p53 dynamics control cell fate. *Science* **336**, 1440–1444 (2012).
- Hao, N., Budnik, B. A., Gunawardena, J. & O’Shea, E. K. Tunable signal processing through modular control of transcription factor translocation. *Science* **339**, 460–464 (2013).
- Young, J. W., Locke, J. C. & Elowitz, M. B. Rate of environmental change determines stress response specificity. *Proc. Natl. Acad. Sci. USA* **110**, 4140–4145 (2013).
- Makman, R. S. & Sutherland, E. W. Adenosine 3',5'-phosphate in *Escherichia coli*. *J. Biol. Chem.* **240**, 1309–1314 (1965).
- Perlmutter, R. L., De Crombrugghe, B. & Pastan, I. Cyclic AMP regulates catabolite and transient repression in *E. coli*. *Nature* **223**, 810–812 (1969).
- Magasanik, B. Catabolite repression. *Cold Spring Harb. Symp. Quant. Biol.* **26**, 249–256 (1961).
- Epps, H. M. & Gale, E. F. The influence of the presence of glucose during growth on the enzymic activities of *Escherichia coli*: comparison of the effect with that produced by fermentation acids. *Biochem. J.* **36**, 619–623 (1942).
- Neidhardt, F. C. & Magasanik, B. Effect of mixtures of substrates on the biosynthesis of inducible enzymes in *Aerobacter aerogenes*. *J. Bacteriol.* **73**, 260–263 (1957).
- Kolb, A., Busby, S., Buc, H., Garges, S. & Adhya, S. Transcriptional regulation by cAMP and its receptor protein. *Annu. Rev. Biochem.* **62**, 749–797 (1993).
- Saier, M. H. Jr, Feucht, B. U. & Hofstadter, L. J. Regulation of carbohydrate uptake and adenylate cyclase activity mediated by the enzymes II of the phosphoenolpyruvate: sugar phosphotransferase system in *Escherichia coli*. *J. Biol. Chem.* **251**, 883–892 (1976).
- Deutscher, J., Francke, C. & Postma, P. W. How phosphotransferase system-related protein phosphorylation regulates carbohydrate metabolism in bacteria. *Microbiol. Mol. Biol. Rev.* **70**, 939–1031 (2006).
- Epstein, W., Rothman-Denes, L. B. & Hesse, J. Adenosine 3':5'-cyclic monophosphate as mediator of catabolite repression in *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **72**, 2300–2304 (1975).
- Hogema, B. M. *et al.* Catabolite repression by glucose 6-phosphate, gluconate and lactose in *Escherichia coli*. *Mol. Microbiol.* **24**, 857–867 (1997).
- Bettenbrock, K. *et al.* Correlation between growth rates, EIICABr phosphorylation, and intracellular cyclic AMP levels in *Escherichia coli* K-12. *J. Bacteriol.* **189**, 6891–6900 (2007).
- Mandelstam, J. The repression of constitutive beta-galactosidase in *Escherichia coli* by glucose and other carbon sources. *Biochem. J.* **82**, 489–493 (1962).
- McFall, E. & Magasanik, B. Effects of thymine and of phosphate deprivation on enzyme synthesis in *Escherichia coli*. *Biochim. Biophys. Acta* **55**, 900–908 (1962).
- Clark, D. J. & Marr, A. G. Studies on the repression of beta-galactosidase in *Escherichia coli*. *Biochim. Biophys. Acta* **92**, 85–94 (1964).
- Ullmann, A., Tillier, F. & Monod, J. Catabolite modulator factor: a possible mediator of catabolite repression in bacteria. *Proc. Natl. Acad. Sci. USA* **73**, 3476–3479 (1976).
- Narang, A. & Pilyugin, S. S. Bacterial gene regulation in diauxic and non-diauxic growth. *J. Theor. Biol.* **244**, 326–348 (2007).
- Monod, J. The phenomenon of enzymatic adaptation - and its bearings on problems of genetics and cellular differentiation. *Growth* **11**, 223–289 (1947).
- Okada, T. *et al.* Role of inducer exclusion in preferential utilization of glucose over melibiose in diauxic growth of *Escherichia coli*. *J. Bacteriol.* **146**, 1030–1037 (1981).
- Inada, T., Kimata, K. & Aiba, H. Mechanism responsible for glucose-lactose diauxie in *Escherichia coli*: challenge to the cAMP model. *Genes Cells* **1**, 293–301 (1996).
- Müller-Hill, B. *The lac Operon: a Short History of a Genetic Paradigm*. (de Gruyter, 1996).
- Kuhlman, T., Zhang, Z., Saier, M. H. Jr & Hwa, T. Combinatorial transcriptional control of the lactose operon of *Escherichia coli*. *Proc. Natl. Acad. Sci. USA* **104**, 6043–6048 (2007).
- Wanner, B. L., Kodaira, R. & Neidhardt, F. C. Regulation of *lac* operon expression: reappraisal of the theory of catabolite repression. *J. Bacteriol.* **136**, 947–954 (1978).
- Kuo, J. T., Chang, Y. J. & Tseng, C. P. Growth rate regulation of *lac* operon expression in *Escherichia coli* is cyclic AMP dependent. *FEBS Lett.* **553**, 397–402 (2003).
- Klumpp, S., Zhang, Z. & Hwa, T. Growth rate-dependent global effects on gene expression in bacteria. *Cell* **139**, 1366–1375 (2009).
- Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. *Science* **330**, 1099–1102 (2010).
- Reitzer, L. Nitrogen assimilation and global regulation in *Escherichia coli*. *Annu. Rev. Microbiol.* **57**, 155–176 (2003).
- Kim, M. *et al.* Need-based activation of ammonium uptake in *Escherichia coli*. *Mol. Syst. Biol.* **8**, 616 (2012).

36. Silverstone, A. E., Arditti, R. R. & Magasanik, B. Catabolite-insensitive revertants of *lac* promoter mutants. *Proc. Natl Acad. Sci. USA* **66**, 773–779 (1970).

37. Schaechter, M., Maaloe, O. & Kjeldgaard, N. O. Dependency on medium and temperature of cell size and chemical composition during balanced growth of *Salmonella typhimurium*. *J. Gen. Microbiol.* **19**, 592–606 (1958).

38. Maaloe, O. in *Biological Regulation and Development* 487–542 (Plenum, 1979).

39. Bender, R. A. & Magasanik, B. Regulatory mutations in the *Klebsiella aerogenes* structural gene for glutamine synthetase. *J. Bacteriol.* **132**, 100–105 (1977).

40. Scott, M. & Hwa, T. Bacterial growth laws and their applications. *Curr. Opin. Biotechnol.* **22**, 559–565 (2011).

41. Schuetz, R., Zamboni, N., Zampieri, M., Heinemann, M. & Sauer, U. Multidimensional optimality of microbial metabolism. *Science* **336**, 601–604 (2012).

42. Goyal, S., Yuan, J., Chen, T., Rabinowitz, J. D. & Wingreen, N. S. Achieving optimal growth through product feedback inhibition in metabolism. *PLOS Comput. Biol.* **6**, e1000802 (2010).

43. Doucette, C. D., Schwab, D. J., Wingreen, N. S. & Rabinowitz, J. D. α -ketoglutarate coordinates carbon and nitrogen utilization via enzyme I inhibition. *Nature Chem. Biol.* **7**, 894–901 (2011).

44. Yan, D., Lenz, P. & Hwa, T. Overcoming fluctuation and leakage problems in the quantification of intracellular 2-oxoglutarate levels in *Escherichia coli*. *Appl. Environ. Microbiol.* **77**, 6763–6771 (2011).

45. Leigh, J. R. *Control Theory* (The Institution of Electrical Engineers, 2004).

46. Peterkofsky, A. & Gazdar, C. *Escherichia coli* adenylate cyclase complex: regulation by the proton electrochemical gradient. *Proc. Natl Acad. Sci. USA* **76**, 1099–1103 (1979).

47. Tyler, B. & Magasanik, B. Physiological basis of transient repression of catabolic enzymes in *Escherichia coli*. *J. Bacteriol.* **102**, 411–422 (1970).

48. Harwood, J. P. & Peterkofsky, A. Glucose-sensitive adenylate cyclase in toluene-treated cells of *Escherichia coli*. *B. J. Biol. Chem.* **250**, 4656–4662 (1975).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to R. Bender, A. Danchin, P. Geiduschek, J. Ingraham, S. Kustu, W. F. Loomis, A. Narang, J. Rabinowitz, M. H. Saier and members of the Hwa laboratory for valuable comments. This work was supported by the Human Frontiers in Science Program (RGP0022), and by the NSF to T.H. (PHY1058793) and through the Center for Theoretical Biological Physics (PHY0822283).

Author Contributions C.Y., D.Y. and T.H. designed the study. C.Y., H.O., S.H., Z.Z. M.K., C.W.G. and D.Y. performed experiments. C.Y., S.H., Y.P.W. and T.H. analysed the data. P.L. and T.H. developed the model. All authors contributed to writing the paper and the supplement.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.H. (hwa@ucsd.edu).

The initiation of mammalian protein synthesis and mRNA scanning mechanism

Ivan B. Lomakin¹ & Thomas A. Steitz^{1,2}

During translation initiation in eukaryotes, the small ribosomal subunit binds messenger RNA at the 5' end and scans in the 5' to 3' direction to locate the initiation codon, form the 80S initiation complex and start protein synthesis. This simple, yet intricate, process is guided by multiple initiation factors. Here we determine the structures of three complexes of the small ribosomal subunit that represent distinct steps in mammalian translation initiation. These structures reveal the locations of eIF1, eIF1A, mRNA and initiator transfer RNA bound to the small ribosomal subunit and provide insights into the details of translation initiation specific to eukaryotes. Conformational changes associated with the captured functional states reveal the dynamics of the interactions in the P site of the ribosome. These results have functional implications for the mechanism of mRNA scanning.

In bacteria, translation initiation is controlled by initiation factors IF1, IF2 and IF3 (ref. 1). In eukaryotes, translation initiation is more complex: multiple initiation factors facilitate the recruitment and scanning of mRNA, selection of initiator tRNA, and the joining of the large and small ribosomal subunits^{2–4}. Eukaryotic initiator tRNA binds specifically to eIF2-GTP, forming a stable ternary complex. The ternary complex forms the 43S pre-initiation complex (PIC) by binding to the small ribosomal subunit (40S). Formation of the 43S PIC is promoted by the eukaryotic initiation factors (eIF) eIF1, eIF1A, eIF3 and eIF5 (ref. 5). 43S PIC then binds to the capped 5' end of mRNA through eIF4F and eIF3, and scans the mRNA up to the first AUG codon.

The scanning hypothesis (first AUG rule) was proposed over 30 years ago⁶ and subsequently refined to reflect that sequences surrounding AUG, particularly at the -3 and +4 positions with respect to the A, are critical for initiation in mammals^{7,8}. Despite widespread acceptance of this model, its molecular mechanism remains unknown^{2,9}. Scanning stops upon AUG recognition, triggering eIF1 dissociation. The release of phosphate from the ternary complex makes this reaction irreversible and promotes dissociation of eIF2-GDP^{10,11}. The 48S PIC that is now formed contains the initiator tRNA base paired with AUG in the P site of the 40S subunit. The subsequent attachment of the large (60S) ribosomal subunit is facilitated by the GTP binding factor eIF5B. eIF5B and eIF1A dissociate from the 40S subunit during or after 80S ribosome assembly, leaving the ribosome primed for elongation².

We have determined the crystal structures of the rabbit 40S subunit in complex with: (1) eIF1 (PIC1); (2) eIF1 and eIF1A (PIC2); and (3) mRNA, tRNA and eIF1A (48S PIC). These structures establish the positions of eIF1, eIF1A, mRNA and tRNA and reveal the architectural principles governing the assembly of the pre-initiation complex on mRNA. Comparison of these structures shows that a conformational rearrangement in the P site and a rotation of the head of the 40S subunit accompany complex assembly. This provides the basis for understanding the functions of eIF1 and eIF1A during scanning.

Crystallization and structure determination

Rabbit 40S ribosomal subunit complexes were crystallized in the $P3_121$ space group and contain one 40S subunit per asymmetric unit (Fig. 1 and Supplementary Fig. 1). Complete data sets were collected to 7.9–9 Å resolution (Supplementary Table 1). The 48S PIC was formed by incubating 40S ribosomal subunits with mRNA, non-aminoacylated,

in vitro transcribed initiator tRNA ($tRNA_i$), and human eIF1A and eIF1. Although the presence of eIF1 in the reaction mixture is essential for crystallization, this factor is not seen in the 48S PIC structure. The PIC1 complex was formed by incubating 40S ribosomal subunits with eIF1, and the PIC2 complex was formed by soaking eIF1A in crystals of PIC1 (Methods).

The presence of additional electron densities in the maps allows us to identify the locations of eIF1, eIF1A, tRNA and mRNA unambiguously (Fig. 1 and Supplementary Fig. 2). We provide *Escherichia coli* nomenclature for ribosomal proteins and rRNA in parentheses. We focus predominantly on the details related to translation initiation, which include the interactions of initiation factors, tRNA and mRNA with the 40S ribosomal subunit and their implications for the mechanistic understanding of the scanning process.

Initiator tRNA in the P site

The P site of the 40S subunit contains a binding pocket, occupied by the anticodon stem-loop (ASL) of $tRNA_i$. This pocket is formed by the 18S rRNA helices h28, h44 of the body, h24 of the platform, h28, h29, h30, h31 of the head domains and by three ribosomal proteins (rpS) located in the head of 40S subunit: rpS15(19), rpS16(9) and rpS18(13) (Fig. 2 and Supplementary Fig. 7). The regions of $tRNA_i$ that interact with rRNA are similar to those seen for the P site tRNA bound to the 70S ribosome^{12,13}. However, the rRNA is arranged differently in the $tRNA_i$ binding pockets of PIC1, PIC2 and 48S PIC, reflecting the differences between the scanning-competent (PIC1, PIC2) and scanning-incompetent (48S PIC) conformations. Movement of the $tRNA_i$ towards the E site is blocked in the 48S PIC by helices h24 and h29, which contact the ASL from the platform and head side of the 40S subunit, respectively (Supplementary Fig. 3). The position of h29 is different in the PIC1 and PIC2 structures compared to that in 48S PIC, and allows the P-site-bound tRNA to move towards the E site (Fig. 2b). The corresponding region of h29 has shifted ~4 Å towards the E site in PIC1 and ~2 Å in PIC2. The binding of $tRNA_i$ on the opposite side of the pocket in the 48S PIC might be stabilized by the carboxy terminus of rpS15(19) and by the amino-terminal tail (NTT) of eIF1A, as shown by the presence of additional electron density near the ASL of the $tRNA_i$ in the corresponding map (Supplementary Fig. 4). The suggested orientation of the C terminus of rpS15(19) in the 48S PIC can explain the crosslink reported between rabbit rpS15(19) and the +4

¹Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, Connecticut 06520-8114, USA. ²Howard Hughes Medical Institute, Yale University, New Haven, Connecticut 06520-8114, USA.

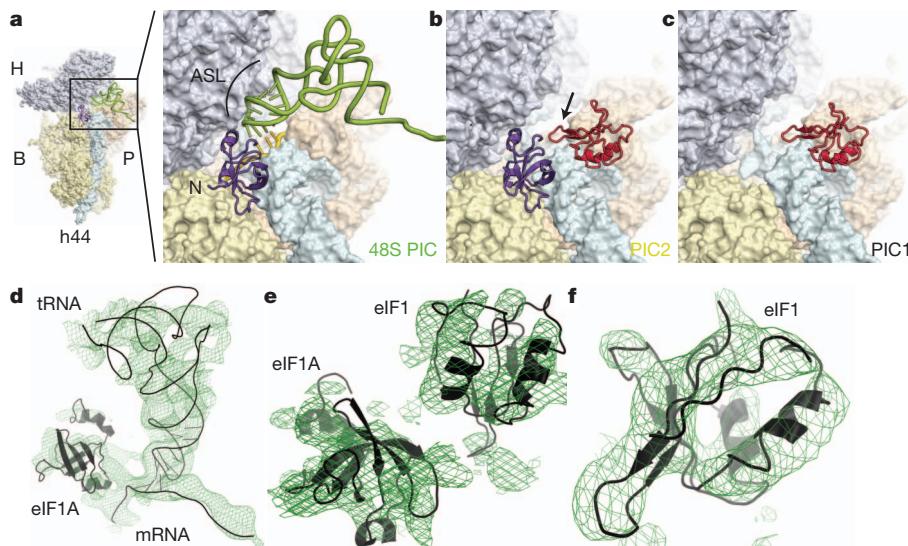


Figure 1 | The crystal structure of rabbit pre-initiation complexes. **a**, 48S PIC. **b**, PIC2. **c**, PIC1. B, body; H, head; h44, helix 44; P, platform. eIF1A (violet), eIF1 (red), tRNA_i (green) and mRNA (yellow) are shown. The basic loop of eIF1 is marked by an arrow; the N terminus of eIF1A is marked by N.

position on mRNA in the 48S complex¹⁴. Positioning of the NTT of eIF1A near the ASL of tRNA_i in 48S PIC is consistent with biochemical data demonstrating that NTT contributes to eIF1A binding to the 40S subunit¹⁵ and may interact with the ASL¹⁶. The NTT of eIF1A remains unstructured in PIC2. We propose that the architecture of the P site in PIC1 and PIC2 represents the scanning-competent conformation of the 40S subunit, with tRNA_i in the P_{out} mode². In contrast, the P site in 48S PIC represents a scanning-incompetent conformation of the 40S subunit, with the ASL of tRNA_i in P_{in} mode: locked in position by base pairing with AUG and interactions with the binding pocket, possibly with rpS15(19) and the NTT of eIF1A (refs 2, 17). This position of tRNA_i is similar to the P/I state in the bacterial 30S initiation complex^{18,19} (Supplementary Fig. 6).

Rearrangements of the P site in the PICs are accompanied by a 3° clockwise rotation of the head of the 40S subunit after the binding of eIF1A to PIC1, and an additional 3° upon formation of the 48S PIC. Rotations of the head shift part of the tRNA_i binding pocket towards the A site by ~2 Å with each rotation. At the same time, h28, which comprises the neck that connects the head and body of the 40S subunit, does not change its position (Fig. 2b and Supplementary Fig. 5).

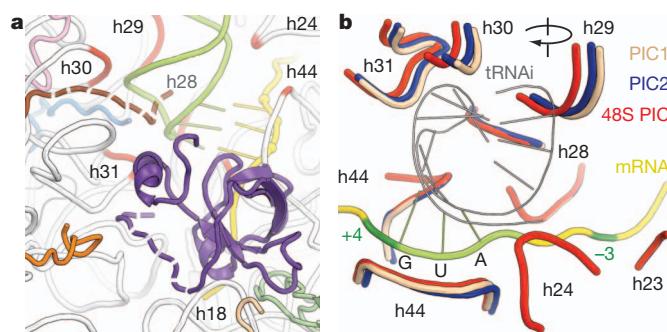


Figure 2 | Interactions of tRNA_i and eIF1A with the 40S subunit. **a**, eIF1A (violet) binds to h44, rpS30e (wheat) and rpS23 (green). Possible positions of the NTT of eIF1A and the C terminus of rpS15 (brown) are shown as dashed lines. Conserved regions of 18S rRNA, which form the tRNA_i binding pocket, are in red. rpS16 (blue), rpS18 (pink) and rpS31 (orange) are shown. **b**, Superposition of 48S PIC, PIC1 and PIC2. Only regions of rRNA that are important for the interaction with the tRNA_i are shown. Clockwise rotation of the head domain is indicated.

d–f, Initial unbiased electron density maps $F_o - F_c$ contoured at $\sigma = 2.0$ (green mesh) for 48S PIC (d), PIC2 (e) and PIC1 (f). Docked structures are shown in black.

The role of h28 as a pivot point for head rotation suggests that mutations in the neck affect rotation of the head and repress the conformational rearrangements of the P site that are required during scanning. This probably explains the leaky scanning phenotype seen in mutational studies of h28 in the yeast 18S rRNA²⁰.

The mRNA channel and the mRNA path

The electron density map of the 48S PIC reveals the position of the mRNA on the 40S subunit from nucleotides −6 to +7 (Fig. 3a). Its conformation does not have the sharp kink between the A and P site codons that is seen in the bacterial 70S ribosomal complexes. This kink allows the simultaneous pairing of the A and P site tRNA anticodons with the mRNA²¹ (Fig. 3a), which defines the reading frame and prevents slippage of the mRNA¹³. The conformation of the mRNA in the scanning complex should promote the slippage of mRNA. eIF1A assists the slippage of the 40S subunit along the mRNA by occupying the A site, which prevents possible interactions between tRNAs and mRNA that might otherwise fix the mRNA on the ribosome (Fig. 3a and Supplementary Fig. 7).

The mRNA path in the 48S PIC is similar to that in 70S ribosomal complexes^{14,22}. Eukaryote-specific features are localized in the mRNA channel near nucleotides −3 and +4, which are the most sensitive positions in the consensus sequence around the initiation codon, and

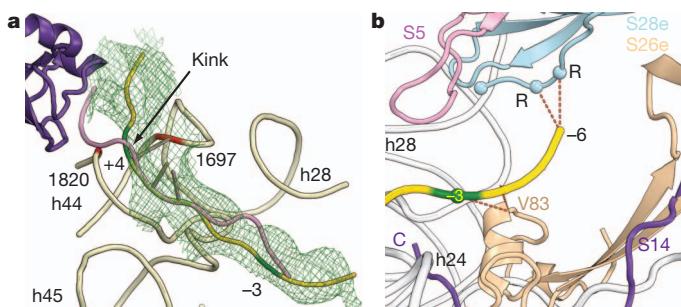


Figure 3 | Path of the mRNA in the 48S PIC. **a**, Initial unbiased electron density map $F_o - F_c$ contoured at $\sigma = 2.0$ (green mesh). eIF1A (violet), mRNA in eukaryotic 48S PIC (yellow) and mRNA from bacterial 70S translating ribosome (pink) are shown. Nucleotides +4 and −3 of the mRNA are in dark green. **b**, rpS26e is ~5 Å away from nucleotide −3 of the mRNA. The C terminus of rpS28e is ~7 Å from the −5, −6 region of the mRNA.

also upstream of the E site, because eukaryotes lack the Shine–Dalgarno sequence (SD)^{22,23}. Two ribosomal proteins, rpS26e and rpS28e, which do not have homologues in eubacteria are present in these regions. rpS26e has an α -helix (C74–V83) that is located \sim 5 Å away from positions –4 and –3 on the mRNA as well as from h28 (Fig. 3b). This helix may change the orientation of the –3 nucleotide during scanning, thus affecting its interaction with the α -subunit of eIF2, implicated in the discrimination of the AUG context at position –3 (ref. 14). Although rpS26e is bound to the platform of the 40S subunit, its N terminus interacts with the neck of the 40S subunit, possibly facilitating communication between position –3 on the mRNA, the head domain and the platform.

mRNA position –6 is located in the vicinity of rpS28e. This region of mRNA is of particular interest because it is involved in the SD interaction in prokaryotes and modulates translation initiation on non-AUG initiation codons in eukaryotes²⁴. The C terminus of rpS28e contains three conserved arginine residues, located \sim 7 Å from nucleotide –6, which may interact with it (Fig. 3b). We suggest that these eukaryote-specific interactions may stabilize the position of mRNA during initiation. The initiation codon and nucleotide +4 are located in a gorge formed by the top of h44 (Fig. 3a and Supplementary Fig. 7). We propose that nucleotides 1697(1401) and 1820(1494) of h44 of 18S rRNA, together with the NTT of eIF1A, modulate the orientation of nucleotide +4 of the mRNA, which may pause mRNA sliding during scanning (Fig. 3a). This pause would allow the AUG codon or ASL of tRNA_i to be accommodated for codon–anticodon base pairing. The presence of electron density attributed to the NTT of eIF1A in the 48S PIC (Supplementary Fig. 4) indicates that upon base pairing with the initiation codon, the NTT might stabilize the P_{in} conformation of the tRNA_i¹⁷ by interacting with the mRNA's +4 nucleotide and/or with the ASL of tRNA_i (Fig. 2 and Supplementary Fig. 7).

Initiation factor eIF1A and the latch

The universally conserved eIF1A stimulates binding of the ternary complex to the 40S subunit and is involved in the scanning and selection of the initiation codon². Its bacterial orthologue IF1 is bound to the A site of the 30S ribosomal subunit, where it sterically prevents the binding of tRNA²⁵. The location of the binding site of eIF1A on the 40S ribosomal subunit was previously mapped to the A site¹⁶. An NMR structure of eIF1A revealed that its main domain has an oligonucleotide/oligosaccharide binding fold that is similar to IF1, an additional C-terminal subdomain with two α -helices, and long unstructured N- and C-terminal tails (CTT)²⁶. Whereas the CTT is required for stringent AUG selection by promoting the scanning-competent conformation of the initiation complex, mutations in the NTT suppress the elevated level of UUG initiation^{15,27}.

Electron density maps of 48S PIC and PIC2 show clear density for eIF1A (Fig. 1)²⁶. Whereas the CTT of eIF1A remains disordered, the NTT is partially visible in the 48S PIC. The additional density in the difference map ($F_o - F_c$) of the 48S PIC can be attributed to the NTT of eIF1A. In this case, it is located 7–8 Å from nucleotide +4 of the mRNA and the ASL of the tRNA_i (Fig. 2 and Supplementary Fig. 4). eIF1A binds to the 40S subunit at the top of h44, which forms a part of the intersubunit bridge B2a (ref. 28), and to h18 of the 18S rRNA through its oligonucleotide/oligosaccharide binding domain, similar to that of IF1 (ref. 25). eIF1A interacts with the N terminus of rpS30e and with rpS23(12) (Fig. 2a and Supplementary Fig. 7). This position of eIF1A ensures that the A site on the 40S subunit remains inaccessible to tRNAs during initiation.

eIF1A forms a bridge between the head and the body of the 40S subunit through its helical subdomain (PIC2 and 48S PIC) and NTT (48S PIC) (Figs 2a and 4). These eukaryote-specific interactions are consistent with hydroxyl radical cleavage experiments¹⁶. Bridging together the head and the body of the 40S subunit ensures the closure of the mRNA channel around the A site, which promotes scanning of the mRNA.

The ‘latch’ of mRNA entry channel remains closed in the structures of all three PICs presented (Fig. 4a). It is formed by interactions between h18 of the body with h34 and ribosomal protein rpS3(3) of the head, which clamp around the incoming mRNA^{29,30}. It was proposed that the closed conformation of the latch is a feature of ribosomes in the active conformation, which “provides a geometry that guarantees processivity and ensures maximum fidelity”³⁰. We propose that closure of the latch is necessary for the scanning function of the 40S ribosomal subunit. Closing of the latch and eIF1A binding alter the environment of the mRNA binding channel and ensure that the mRNA remains unstructured, correctly oriented for scanning and free to move towards the exit.

The closed conformation of the latch in our structure of PIC2 is different from the open conformation of the latch in the cryo-electron microscopy reconstruction of yeast 40S subunit complex with eIF1 and eIF1A (ref. 29). It is possible that eIF3 may be involved in the latch opening in mammals, because unlike in yeast, eIF3 is required for ternary complex to bind to the 40S subunit. The latch closure might be influenced by crystallization, although the crystal packing does not affect rotation of the head domain of the 40S subunit. Furthermore, the latch could sample different conformational states during initiation. The open conformation of the latch allows direct loading of the mRNA in the mRNA binding channel, instead of threading it through a tunnel²⁹. We propose that the cryo-electron microscopy structure of the yeast 40S subunit with bound eIF1 and eIF1A represents the mRNA binding state of the ribosome, whereas the rabbit PIC2 represents its scanning mode.

eIF1 is poised to destabilize tRNA_i binding

eIF1 is a functional homologue of the bacterial IF3 (refs 31, 32). During scanning, eIF1 allows proofreading and sensing a non-optimal context of AUG codons, while promoting the dissociation of aberrantly assembled complexes^{33,34}. eIF1 binds to the top of h44 and to helices h24 and h45 of the 18S rRNA in the vicinity of the P site (Figs 1 and 4b), consistent with previously published studies^{31,35,36}. The region of interaction between eIF1 and 18S rRNA consists of the intersubunit bridge B2a and part of the bridge B2b (ref. 28). Superposition of the 48S PIC onto PIC1 and PIC2 reveals moderate sterical clashes between eIF1 and tRNA_i, as was predicted earlier by comparison of structures of 40S subunit complex with eIF1 from *Tetrahymena thermophila* and bacterial 70S ribosomal complexes³⁵. The basic loop of eIF1 (R38–K42) in both PIC1 and PIC2 is positioned between the ASL of tRNA_i and nucleotides 1821–1823(1495–1497) of h44, which stabilizes the interaction between AUG and the anticodon in the P site of the 48S PIC. This results in sterical clashes between the basic loop of eIF1 and the ASL of tRNA_i and between the region around residues P77–G80 of eIF1 and the D stem of tRNA_i (Fig. 4b).

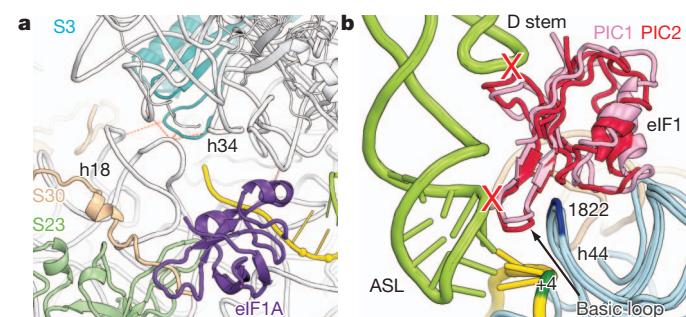


Figure 4 | The latch and initiation factor eIF1. **a**, Closed conformation of the latch. Interactions between body and head domains are connected by dashed lines. **b**, The P site of the 40S subunit. eIF1 from PIC1 (pink) or PIC2 (red) would sterically clash (red X) with the ASL and the D stem of tRNA_i (green) that is positioned in the 48S PIC.

It has been proposed that eIF1 modulates recognition of the start codon by restricting tRNA_i binding to AUG only in an optimal context^{33,34}. Fluorescent resonance energy transfer experiments (FRET) with eIF1 and eIF1A demonstrated that AUG recognition triggers the dissociation of eIF1 from 48S PIC, which allows the release of the phosphate (P_i) from the eIF2–GDP–P_i complex^{10,11}. Our data provide an explanation by revealing that eIF1 spatially interferes with the adjustment of the ASL of the initiator tRNA in the P site during scanning. Establishing the interaction between the AUG codon and the anticodon results in locking of the P site tRNA, which can prevent the basic loop of eIF1 from re-entering the P site. Kinetic experiments show that the dissociation constant of eIF1 from the 40S subunit is around 1.4 nM in the presence of the ternary complex, whereas the high affinity of the ternary complex to the 40S subunit results from base pairing with mRNA^{11,29}. This affinity is low when no mRNA is bound to the 40S subunit or when only one nucleotide in the P site is base paired. It is increased with the ability of the initiator tRNA to base pair with two nucleotides and becomes very strong upon base pairing with the AUG (dissociation constant (K_d) < 1 nM)^{11,37}. At this point, upon the correct codon–anticodon base pairing eIF1 cannot compete with the ASL of tRNA_i for the P site and must dissociate.

The basic loop of eIF1 is located within 5 Å of h44 and 10 Å of the +4 position on mRNA (Fig. 4b). That may allow eIF1 to ‘inspect’ the important region downstream of AUG. The same region of h44 undergoes small conformational changes upon formation of 48S PIC, which can weaken its interaction with eIF1 and contribute to the dissociation of eIF1 from the 40S subunit. Consequently, our structures of rabbit PICs provide mechanistic insights into the current model of eukaryotic translation initiation.

Scanning model

The observed conformations of the pre-initiation complexes described in this study together with previous biochemical data² allow us to propose a model for scanning. Briefly, the mRNA slides through the tunnel, formed by the body and the head of the 40S subunit (the latch), eIF1A and eIF1. The tunnel ensures scanning processivity by keeping the mRNA unstructured and properly oriented for the ‘inspection’ of the nucleotide sequence in the P site by tRNA_i. The tRNA_i ‘inspects’ or scans the mRNA by attempting to establish Watson–Crick base pairing between its anticodon and a nucleotide triplet of mRNA moving through the P site. The tRNA_i is in the P_{out} state because the orientation of the head domain of the 40S subunit allows a minor movement of the tRNA_i to avoid a steric clash with eIF1. The affinity of the basic loop of eIF1 to the P site is sufficient to displace the ASL of tRNA_i and prevent the locking of the tRNA_i in the P site (Fig. 5).

When the AUG codon is in the P site it becomes base paired with all three nucleotides of the anticodon of the tRNA_i, thereby stabilizing the conformation of the tRNA_i while allowing it to displace the basic loop of eIF1. Subsequent rotation of the head domain locks tRNA_i in the P_{in} state and completes the scanning (Fig. 5). The transition from P_{out} to P_{in} mode of tRNA_i is probably promoted by the C terminus of rpS15(19) and the NTT of eIF1A (ref. 17), which together might

stabilize the base-paired conformation of tRNA_i in the P site. In addition, a similar effect may be caused by the interactions of the nucleotide +4 on the mRNA with h44, eIF1A and eIF1, and the nucleotide –3 with h23, rpS5(7) and rpS26e. The dissociation of eIF1 and eIF2 from the 40S subunit results in a stable 48S PIC similar to the one observed in this study. eIF5B then binds to this complex and orients the acceptor stem of tRNA_i towards the P site on the 60S subunit to facilitate subunit joining in a manner that is similar to its bacterial analogue IF2 (Supplementary Fig. 6).

Conformational changes in the P site may affect the interaction of eIF1 with the top of h44 to which it binds. This probably provides synchronization between the rotations of the head domain of the 40S subunit, which rearrange the P site, and the presence of eIF1 on the ribosome. Interaction of eIF1A with h44 may cause the same effect and explain the cooperative binding of eIF1A and eIF1 to the 40S subunit^{29,38}. Future high-resolution structures of the 40S ribosomal subunit complexes with the initiation factors, mRNA and tRNA, coupled with biochemical studies, are needed to refine and complete the model presented here.

METHODS SUMMARY

40S ribosomal subunits were purified from rabbit reticulocyte lysate. Phases for calculating the electron density maps were determined by molecular replacement using as a search model the structure of the 40S subunit from the human 80S ribosome (Supplementary Tables 1–3 and Methods³⁹). Modelling of additional rRNA and protein regions was impossible due to restrictions imposed by the resolution limits.

Full Methods and any associated references are available in the online version of the paper.

Received 11 March; accepted 7 June 2013.

Published online 21 July 2013.

1. Myasnikov, A. G., Simonetti, A., Marzi, S. & Klaholz, B. P. Structure-function insights into prokaryotic and eukaryotic translation initiation. *Curr. Opin. Struct. Biol.* **19**, 300–309 (2009).
2. Hinnebusch, A. G. Molecular mechanism of scanning and start codon selection in eukaryotes. *Microbiol. Mol. Biol. Rev.* **75**, 434–467 (2011).
3. Aitken, C. E. & Lorsch, J. R. A mechanistic overview of translation initiation in eukaryotes. *Nature Struct. Mol. Biol.* **19**, 568–576 (2012).
4. Voigt-Hoffmann, F., Klinge, S. & Ban, N. Structural insights into eukaryotic ribosomes and the initiation of translation. *Curr. Opin. Struct. Biol.* **22**, 768–777 (2012).
5. Lorsch, J. R. & Dever, T. E. Molecular view of 43 S complex formation and start site selection in eukaryotic translation initiation. *J. Biol. Chem.* **285**, 21203–21207 (2010).
6. Kozak, M. How do eukaryotic ribosomes select initiation regions in messenger RNA? *Cell* **15**, 1109–1123 (1978).
7. Cavener, D. R. & Ray, S. C. Eukaryotic start and stop translation sites. *Nucleic Acids Res.* **19**, 3185–3192 (1991).
8. Kozak, M. An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs. *Nucleic Acids Res.* **15**, 8125–8148 (1987).
9. Jackson, R. J., Hellen, C. U. & Pestova, T. V. The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Rev. Mol. Cell Biol.* **11**, 113–127 (2010).
10. Algire, M. A., Maag, D. & Lorsch, J. R. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* **20**, 251–262 (2005).
11. Maag, D., Fekete, C. A., Gryczynski, Z. & Lorsch, J. R. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell* **17**, 265–275 (2005).
12. Noller, H. F., Hoang, L. & Fredrick, K. The 30S ribosomal P site: a function of 16S rRNA. *FEBS Lett.* **579**, 855–858 (2005).
13. Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
14. Pisarev, A. V. et al. Specific functional interactions of nucleotides at key –3 and +4 positions flanking the initiation codon with components of the mammalian 48S translation initiation complex. *Genes Dev.* **20**, 624–636 (2006).
15. Fekete, C. A. et al. N- and C-terminal residues of eIF1A have opposing effects on the fidelity of start codon selection. *EMBO J.* **26**, 1602–1614 (2007).
16. Yu, Y. et al. Position of eukaryotic translation initiation factor eIF1A on the 40S ribosomal subunit mapped by directed hydroxyl radical probing. *Nucleic Acids Res.* **37**, 5167–5182 (2009).
17. Saini, A. K., Nanda, J. S., Lorsch, J. R. & Hinnebusch, A. G. Regulatory elements in eIF1A control the fidelity of start codon selection by modulating tRNA(i)(Met) binding to the ribosome. *Genes Dev.* **24**, 97–110 (2010).

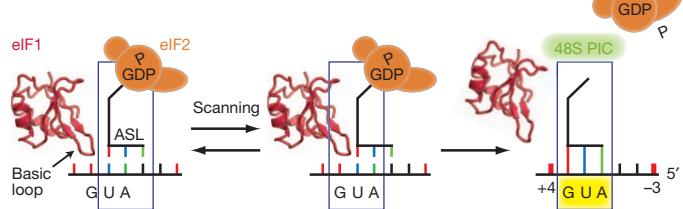


Figure 5 | Scanning model. The basic loop of eIF1 competes for the P site of the 40S subunit with the ASL of tRNA_i, which results in displacement of each other during mRNA scanning.

18. Julián, P. et al. The Cryo-EM structure of a complete 30S translation initiation complex from *Escherichia coli*. *PLoS Biol.* **9**, e1001095 (2011).
19. Simonetti, A. et al. Structure of the 30S translation initiation complex. *Nature* **455**, 416–420 (2008).
20. Dong, J. et al. Genetic identification of yeast 18S rRNA residues required for efficient recruitment of initiator tRNA(Met) and AUG selection. *Genes Dev.* **22**, 2242–2255 (2008).
21. Yusupov, M. M. et al. Crystal structure of the ribosome at 5.5 Å resolution. *Science* **292**, 883–896 (2001).
22. Pisarev, A. V., Kolupaeva, V. G., Yusupov, M. M., Hellen, C. U. & Pestova, T. V. Ribosomal position and contacts of mRNA in eukaryotic translation initiation complexes. *EMBO J.* **27**, 1609–1621 (2008).
23. Kozak, M. Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* **44**, 283–292 (1986).
24. Wegrzyn, J. L., Drudge, T. M., Valafar, F. & Hook, V. Bioinformatic analyses of mammalian 5'-UTR sequence properties of mRNAs predicts alternative translation initiation sites. *BMC Bioinformatics* **9**, 232 (2008).
25. Carter, A. P. et al. Crystal structure of an initiation factor bound to the 30S ribosomal subunit. *Science* **291**, 498–501 (2001).
26. Battiste, J. L., Pestova, T. V., Hellen, C. U. & Wagner, G. The eIF1A solution structure reveals a large RNA-binding surface important for scanning function. *Mol. Cell* **5**, 109–119 (2000).
27. Fekete, C. A. et al. The eIF1A C-terminal domain promotes initiation complex assembly, scanning and AUG selection *in vivo*. *EMBO J.* **24**, 3588–3601 (2005).
28. Ben-Shem, A. et al. The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* **334**, 1524–1529 (2011).
29. Passmore, L. A. et al. The eukaryotic translation initiation factors eIF1 and eIF1A induce an open conformation of the 40S ribosome. *Mol. Cell* **26**, 41–50 (2007).
30. Schluenzen, F. et al. Structure of functionally activated small ribosomal subunit at 3.3 Å resolution. *Cell* **102**, 615–623 (2000).
31. Lomakin, I. B., Kolupaeva, V. G., Marintchev, A., Wagner, G. & Pestova, T. V. Position of eukaryotic initiation factor eIF1 on the 40S ribosomal subunit determined by directed hydroxyl radical probing. *Genes Dev.* **17**, 2786–2797 (2003).
32. Lomakin, I. B., Shirokikh, N. E., Yusupov, M. M., Hellen, C. U. & Pestova, T. V. The fidelity of translation initiation: reciprocal activities of eIF1, IF3 and YciH. *EMBO J.* **25**, 196–210 (2006).
33. Pestova, T. V., Borukhov, S. I. & Hellen, C. U. Eukaryotic ribosomes require initiation factors 1 and 1A to locate initiation codons. *Nature* **394**, 854–859 (1998).
34. Pestova, T. V. & Kolupaeva, V. G. The roles of individual eukaryotic translation initiation factors in ribosomal scanning and initiation codon selection. *Genes Dev.* **16**, 2906–2922 (2002).
35. Rabl, J., Leibundgut, M., Ataide, S. F., Haag, A. & Ban, N. Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* **331**, 730–736 (2011).
36. Fletcher, C. M., Pestova, T. V., Hellen, C. U. & Wagner, G. Structure and interactions of the translation initiation factor eIF1. *EMBO J.* **18**, 2631–2637 (1999).
37. Koltz, S. E., Takacs, J. E. & Lorsch, J. R. Kinetic and thermodynamic analysis of the role of start codon/anticodon base pairing during eukaryotic translation initiation. *RNA* **15**, 138–152 (2009).
38. Maag, D. & Lorsch, J. R. Communication between eukaryotic translation initiation factors 1 and 1A on the yeast small ribosomal subunit. *J. Mol. Biol.* **330**, 917–924 (2003).
39. Anger, A. M. et al. Structures of the human and *Drosophila* 80S ribosome. *Nature* **497**, 80–85 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the members of the Steitz laboratory for useful suggestions and discussions, the staff of the Advanced Photon Source beamline 24-ID, the National Synchrotron Light Source beamlines X25 and X29, and the Richards Center facility at Yale University for support. We also thank Y. Polikanov, J. Wang and Y. Xiong for advice with crystallographic software. This work was supported by the National Institutes of Health (NIH) grant GM022778 (to T.A.S.).

Author Contributions I.B.L. designed and performed experiments, analysed data and wrote the paper; T.A.S. analysed data, wrote the paper and directed research.

Author Information The structural coordinates of PIC1, PIC2 and 48S PIC have been deposited in the Protein Data Bank (<http://www.rcsb.org/pdb>) under accession codes 4KZX, 4KZY and 4KZZ, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.A.S. (thomas.steitz@yale.edu) or I.B.L. (ivan.lomakin@yale.edu).

METHODS

Cloning and protein expression. The DNAs encoding full-length human eIF1 and eIF1A with additional N-terminal affinity tag and TEV protease cleavage site (↓) sequences (MGHHHHHHHDYDIQTTENLYFQ↓G) were synthesized by PCR and cloned into pET28b expression vector (Novagen). Recombinant initiation factors were expressed in *E. coli* Rosetta2(DE3)pLysS cells (Novagen).

Protein purification. Recombinant eIF1 and eIF1A were purified using a HisTrap HP column (GE Healthcare), essentially as described⁴⁰. The proteins were then treated with ActEV protease (Invitrogen) according to the manufacturer's instructions. The affinity tag was removed by chromatography on a HisTrap HP column (GE Healthcare). Tag-free eIF1 and eIF1A were further purified on Mono S and Mono Q columns (GE Healthcare), respectively. Finally, both proteins were purified on a Superdex 75 column (GE Healthcare) equilibrated with buffer H300 (5 mM HEPES-KOH, 300 mM KCl, 5% glycerol, pH 7.1). eIF1 and eIF1A were frozen in liquid nitrogen in small aliquots and stored at -80 °C.

Initiator tRNA. Mammalian initiator tRNA⁴¹ was transcribed using T7 RNA polymerase and chemically synthesized DNA oligonucleotides as suggested in ref. 42. tRNA was extracted by a phenol chloroform mixture (5:1, pH 4.3) and precipitated by ethanol. tRNA was dissolved in buffer (10 mM Tris-HCl, 90 mM KCl, 3.0 mM Mg acetate, 0.5 mM EDTA, pH 7.4) and purified on a Superdex 75 column (GE Healthcare), equilibrated with the same buffer. tRNA was not aminoacylated. tRNA was concentrated to 2.9 mg ml⁻¹, frozen in liquid nitrogen in small aliquots and stored at -80 °C.

mRNA. A 24-nucleotide-long mRNA with an AUG start codon was chemically synthesized by Thermo Fisher Scientific; the sequence was 5'-CAACAAACAACAACAAAUGUUUCAA-3'.

Purification of 40S ribosomal subunits from rabbit reticulocyte lysate (RRL). Frozen RRL was purchased from Green Hectares. The following components were added to 1 l of thawed RRL: 5 mM HEPES-KOH (pH 7.1), 15 mM KCl, 11 mM Mg acetate, 1.0 mM EDTA (pH 8.0), 2 mM DTT, 10 µM MG132, 500 mg heparin, 150 mg Pefabloc, 12 tablets of Complete Protease Inhibitor Cocktail (Roche). Ribosomes were pelleted by centrifugation at 40,000 r.p.m. for 10.5 h in a Beckman Type 45 Ti rotor at 4 °C through a sucrose cushion (7 ml/tube; 30% (w/v) sucrose, 20 mM Bis-Tris (pH 5.9), 200 mM NH₄ acetate, 300 mM KCl, 10 mM Mg acetate, 5 mM DTT). The pellet was re-suspended in 25 ml of buffer A (20 mM HEPES-KOH, 30 mM KCl, 11 mM Mg acetate, 1 mM EDTA, 2 mM DTT, 10 µM MG132, 1.0 mg ml⁻¹ heparin, 0.2 mg ml⁻¹ Pefabloc, 1 tablet/100 ml Complete Protease Inhibitor Cocktail (Roche), 200 U ml⁻¹ SUPERaseIN (Ambion), pH 7.1). Subsequently, NH₄Cl was added to the suspension to a final concentration of 0.4 M and the suspension was mixed on ice for at least 0.3 h. One-third of the suspension was then clarified by centrifugation in an Eppendorf microcentrifuge for 10 min at 13,200 r.p.m. at 4 °C and loaded on a Sephadryl S-400 HR column (GE Healthcare) equilibrated with buffer B (20 mM Tris-HCl, 500 mM KCl, 10 mM MgCl₂, 5% (w/v) glycerol, 2 mM DTT, pH 7.4). This step was repeated for the rest of the ribosomal suspension. Fractions (8 ml) were collected and analysed by SDS-PAGE (4–12%, MES SDS running buffer, NuPAGE). Those containing 80S ribosomes were combined, brought to a final concentration of 2.5 mM MgCl₂ and 1% glycerol and concentrated to a final volume of 25 ml. Optical density of concentrated ribosomes (OD_{260 nm}) should be 200–350 U ml⁻¹. Ribosomes were subsequently supplemented with 1 mM puromycin and incubated for 20 min at 37 °C followed by 20 min incubation at room temperature and then chilled on ice for at least 10 min.

Chilled ribosomes were clarified by centrifugation in an Eppendorf microcentrifuge for 10 min at 13,200 r.p.m. at 4 °C and loaded on a 10–40% sucrose density gradient, prepared in tubes for Beckman SW 32 Ti rotor (10–40% (w/v) sucrose, 20 mM Tris-HCl, 500 mM KCl, 5 mM Mg acetate, 2 mM DTT, pH 7.4). Sucrose density gradient centrifugation was performed at 25,000 r.p.m. for 14.5 h in a Beckman SW 32 Ti rotor. Fractions (1 ml) were collected and analysed by SDS-PAGE (4–12%, MES SDS running buffer, NuPAGE). Fractions containing 40S subunits were combined, diluted five times with buffer B without KCl and loaded on a DEAE column equilibrated in buffer B with 50 mM KCl. Elution was performed by a 0.05–1.0 M KCl gradient in buffer B. Fractions containing 40S subunits were combined, concentrated to 391 U ml⁻¹ (OD_{260 nm}) and dialysed against buffer C (10 mM HEPES-KOH, 30 mM KCl, 5 mM Mg acetate, 5% (w/v) glycerol, 2 mM DTT, pH 7.1). Small aliquots were frozen in liquid nitrogen and stored at -80 °C.

Activity test. Purified 40S ribosomal subunits were tested by toeprinting analysis for their ability to form initiation complexes with mRNA containing CrPV IRES, as described previously⁴³.

Complex formation. Before crystallization, PIC1 complex was formed by mixing ribosomes and eIF1 to final concentrations of 2.6 µM and 9.1 µM, respectively, with subsequent dialysis against buffer DB (2.0 mM HEPES-KOH, 100 mM NH₄Cl, 15 mM KCl, 3 mM Mg acetate, 2 mM TCEP, pH 7.1) for 2 h at room temperature.

The complex was incubated at 37 °C for 15 min and kept at room temperature for crystallization setup. Complex PIC2 was formed by soaking eIF1A (final concentration 40 µM) during freezing crystals of the PIC1. The 48S PIC was formed similarly to PIC1 but with the addition of eIF1A (final concentration 7.5 µM), mRNA (10 µM) and tRNA_i (10 µM). Although the presence of eIF1 in the reaction mixture is essential for crystallization, this factor is not seen in the 48S PIC structure. We suggest that it promotes the homogeneity of the complex, which is required for crystal growth, by controlling the correct recognition of the initiation codon by tRNA_i and by dissociating aberrantly assembled ribosomal complexes^{33,34}.

Crystallization and cryoprotection. Crystals were grown in 24-well sitting-drop plates using the vapour diffusion technique. Four microlitres of the PIC1 were mixed with 4 µl of reservoir solution (50 mM Tris-HCl, 100 mM NH₄ acetate, 5 mM Mg acetate, 50 mM sarcosine, 5 mM urea, 2.5% PEG 20000, 5% MPD, 2 mM TCEP, pH 8.5). Plates were incubated at 19 °C for 6–9 days. Crystals were stabilized by stepwise increase of MPD to the final concentration of 35%. After stabilization, crystals were frozen in a cryo stream. For PIC2 complex formation, crystals were stabilized in the same buffer containing 40 µM of eIF1A.

48S PIC (4 µl) was mixed with 7.4 µl of reservoir solution (50 mM Tris-HCl, 100 mM NH₄ acetate, 5 mM Mg acetate, 50 µM Zn acetate, 2 mM putrescine, 2.5% PEG 20000, 5% 1,4-butanediol, 2 mM TCEP, pH 8.5). Plates were incubated at 19 °C for 6–13 days. Crystals were stabilized at 4 °C by stepwise increase of 1,4-butanediol and PEG 20000 to the final concentration of 30% and 3.0%, respectively. After stabilization, crystals were frozen in liquid nitrogen.

Data collection and processing. Data collection was carried out at 100 K at beamlines X25 and X29 at Brookhaven National Laboratory and 24ID at the Advanced Photon Source. A complete data set was collected from a single crystal (PIC1, PIC2) or from two crystals (48S PIC, Supplementary Table 1). Data were integrated and scaled with the XDS program package⁴⁴. All crystals belong to the hexagonal space group *P*3₁21 (Supplementary Table 1).

The structure was solved by molecular replacement using PHASER from the CCP4 program suite⁴⁵. The search model was generated from the previously published structure of yeast 80S ribosome (ref. 28, Protein Data Bank entries 3U5B and 3U5C). Both the 60S subunit and the suppressor protein STM1 were excluded from the search model. Before refinement all B factors in the model were set to an isotropic B of 80. The initial molecular replacement solution was refined by rigid-body refinement with the 40S subunit split into multiple domains (Supplementary Table 2), followed by five cycles of grouped TLS and grouped B-factor refinement using PHENIX⁴⁶. Bulk solvent correction was applied as recommended⁴⁷. After initial refinement, the difference electron density corresponding to each initiation factor, mRNA and tRNA became clearly visible in the difference electron density map ($F_o - F_c$, Fig. 1). The NMR structures of human eIF1 and eIF1A (refs 36, 26, PDB entries 2IF1 and 1D7Q) without their flexible tails were docked into the corresponding electron density, followed by refinement using PHENIX⁴⁶. Structures of the mRNA and P site tRNA were taken from the previously published structure of the 70S ribosome with bound mRNA, and tRNAs (ref. 13, PDB entry 2J00) were docked in the corresponding electron density in 48S PIC map and refined as above. The position of the mRNA (except AUG) was manually adjusted to fit the electron density before their refinement in PHENIX⁴⁶. mRNA and tRNA were treated as one ridged body during refinement.

Despite that at a resolution of about 8–11 Å the structures cannot be built *de novo*, the availability of the high-resolution models allows the detection of the movement of the structural domains of the ribosome, the location of positions of the bound protein factors and even antibiotics^{48–50}. The conformational changes of the 40S subunit described in our manuscript are reproducible during refinement independently of the starting model: for example, refining the 40S structure from PIC1 against the data of PIC2 results in conformation of PIC2, whereas refining the 40S structure from PIC2 against the data of PIC1 results in conformation of PIC1. That assures us that movements of the domains of the 40S subunit that were observed are not biased to the starting models.

When the manuscript was under review the high-resolution cryo-electron microscopy structure of the human 80S ribosome was published³⁹. Because of the high identity between rabbit and human ribosomal proteins and 18S rRNA (Supplementary Table 3) we used the structure of the human 40S subunit for the final refinement.

Figures. Figures showing electron densities and atomic models were generated using PYMOL (The PyMOL Molecular Graphics System, <http://www.pymol.org>).

40. Pestova, T. V., Hellen, C. U. & Shatsky, I. N. Canonical eukaryotic initiation factors determine initiation of translation by internal ribosomal entry. *Mol. Cell. Biol.* **16**, 6859–6869 (1996).
41. Pestova, T. V. & Hellen, C. U. Preparation and activity of synthetic unmodified mammalian tRNA(Met) in initiation of translation *in vitro*. *RNA* **7**, 1496–1505 (2001).

42. Acker, M. G., Kolitz, S. E., Mitchell, S. F., Nanda, J. S. & Lorsch, J. R. Reconstitution of yeast translation initiation. *Methods Enzymol.* **430**, 111–145 (2007).
43. Pestova, T. V., Lomakin, I. B. & Hellen, C. U. Position of the CrPV IRES on the 40S subunit and factor dependence of IRES/80S ribosome assembly. *EMBO Rep.* **5**, 906–913 (2004).
44. Kabsch, W. Xds. *Acta Crystallogr. D* **66**, 125–132 (2010).
45. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
46. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
47. Rees, B., Jenner, L. & Yusupov, M. Bulk-solvent correction in large macromolecular structures. *Acta Crystallogr. D* **61**, 1299–1301 (2005).
48. Vila-Sanjurjo, A. *et al.* X-ray crystal structures of the WT and a hyper-accurate ribosome from *Escherichia coli*. *Proc. Natl Acad. Sci. USA* **100**, 8682–8687 (2003).
49. Lancaster, L. *et al.* The location of protein S8 and surrounding elements of 16S rRNA in the 70S ribosome from combined use of directed hydroxyl radical probing and X-ray crystallography. *RNA* **6**, 717–729 (2000).
50. Vila-Sanjurjo, A., Schuwirth, B. S., Hau, C. W. & Cate, J. H. Structural basis for the control of translation initiation during stress. *Nature Struct. Mol. Biol.* **11**, 1054–1059 (2004).

A variable absorption feature in the X-ray spectrum of a magnetar

Andrea Tiengo^{1,2,3}, Paolo Esposito², Sandro Mereghetti², Roberto Turolla^{4,5}, Luciano Nobili⁴, Fabio Gastaldello², Diego Götz⁶, Gian Luca Israel⁷, Nanda Rea⁸, Luigi Stella⁷, Silvia Zane⁵ & Giovanni F. Bignami^{1,2}

Soft- γ -ray repeaters (SGRs) and anomalous X-ray pulsars (AXPs) are slowly rotating, isolated neutron stars that sporadically undergo episodes of long-term flux enhancement (outbursts) generally accompanied by the emission of short bursts of hard X-rays^{1,2}. This behaviour can be understood in the magnetar model^{3–5}, according to which these sources are mainly powered by their own magnetic energy. This is supported by the fact that the magnetic fields inferred from several observed properties^{6–8} of SGRs and AXPs are greater than—or at the high end of the range of—those of radio pulsars. In the peculiar case of SGR 0418+5729, a weak dipole magnetic moment is derived from its timing parameters⁹, whereas a strong field has been proposed to reside in the stellar interior^{10,11} and in multipole components on the surface¹². Here we show that the X-ray spectrum of SGR 0418+5729 has an absorption line, the properties of which depend strongly on the star's rotational phase. This line is interpreted as a proton cyclotron feature and its energy implies a magnetic field ranging from 2×10^{14} gauss to more than 10^{15} gauss.

On 2009 June 5 two short bursts of hard X-rays, detected by Fermi and other satellites, revealed the previously unknown source SGR 0418+5729¹³. Subsequent observations with the Rossi X-ray Timing Explorer (RXTE), Swift, Chandra and X-ray Multi-mirror Mission (XMM) Newton satellites found the new SGR to be an X-ray pulsar with a period of ~ 9.1 s and a luminosity of $\sim 1.6 \times 10^{34}$ erg s⁻¹ (in the 0.5–10 keV band and for a distance of 2 kpc)^{13,14}. During the three years after the onset of the outburst, the spectrum softened and the luminosity declined by three orders of magnitude, but remained still too high to be powered by rotational energy^{9,10,14}. The measured spin-down rate of 4×10^{-15} s s⁻¹ translates (under the assumption of a rotating magnetic dipole *in vacuo*) into a magnetic field $B = 6 \times 10^{12}$ G at the magnetic equator⁹, a value well in the range of normal radio pulsars. However, the presence of high-order multipolar field components of 10^{14} G close to the surface has been invoked to interpret the spectrum of the source in the framework of atmosphere models¹². In any case, a strong crustal magnetic field ($> 10^{14}$ G) seems to be required to explain the overall properties of SGR 0418+5729 within the magnetar model^{9,11}.

Hints of the presence of an absorption feature at 2 keV in the spectrum of SGR 0418+5729 were found in the phase-resolved analysis of data (with relatively low-count statistics) from the Swift X-ray Telescope (XRT) taken during 2009 July 12–16¹⁴. Thanks to the large collecting area and good spectral resolution of the European Photon Imaging Camera (EPIC), we were able to perform a more detailed investigation using data collected by XMM-Newton during a 67-ks long observation performed on 2009 August 12, when the source flux was still high (5×10^{-12} erg cm⁻² s⁻¹ in the 2–10 keV band).

To examine the spectral variations as a function of the star's rotational phase without making assumptions about the X-ray spectral energy distribution of SGR 0418+5729, we produced a phase–energy image by binning the EPIC source counts into energy and rotational

phase channels and then normalizing to the phase-averaged energy spectrum and pulse profile. The normalized phase–energy image (Fig. 1) shows a prominent V-shaped feature in the phase interval ~ 0.1 –0.3. This is produced by a lack of counts in a narrow energy range with respect to nearby energy channels, that is, an absorption feature at a phase-dependent energy. The regular shape of the feature in the phase–energy plane as well as its presence in the three independent EPIC detectors (see Supplementary Fig. 5) exclude the possibility that it results from statistical fluctuations in the number of counts or from an instrumental effect. Another absorption feature is visible at low energies at phase ~ 0.5 –0.6.

We extracted from the EPIC data the phase-averaged spectrum of SGR 0418+5729, as well as the spectra from 50 phase intervals of width 0.02 rotational cycles, as described in the Supplementary Information. The phase-averaged spectrum can be adequately fitted by either a two-blackbody model ($\chi^2_v = 1.198$ for 196 degrees of freedom, d.f.) or a blackbody plus power-law model ($\chi^2_v = 1.105$ for 196 d.f.), corrected for interstellar absorption (see refs 11 and 12 for other models that can fit the spectrum).

The 15 spectra extracted from the phase intervals 0.1–0.3 and 0.5–0.6, unlike those of the remaining phases, cannot be fitted by a renormalization of the phase-averaged best-fit model, which gives in most cases null hypothesis probabilities in the range 10^{-4} – 10^{-9} (see Supplementary Fig. 4). They are instead well fitted (null hypothesis probability > 0.03) by the addition of a narrow absorption line component, which can be equally well modelled with a Gaussian profile or a cyclotron absorption line model¹⁵ (the improvement obtained by adding a cyclotron component in the phase intervals 0.1–0.3 and 0.5–0.6 can be seen in Supplementary Fig. 4). The best-fit line parameters as a function of phase are shown in Fig. 2 and an example of a phase-resolved spectrum is displayed in Fig. 3.

We searched for the phase-dependent absorption feature in all the available X-ray observations of SGR 0418+5729 and found that it was present in the phase interval 0–0.3, and up to higher energies than in XMM-Newton, in RXTE data taken during the first two months of the outburst (see Supplementary Fig. 6).

Absorption features have been observed in the X-ray spectra of various classes of neutron stars^{16–23} and interpreted as being due to either cyclotron absorption (by electrons or protons) or bound–bound atomic transitions. However, variations in the line energy as a function of the rotational phase as large as in SGR 0418+5729 (by a factor of ~ 5 in one-tenth of a cycle) have not been seen in any source.

In a neutron star atmosphere, different atomic transitions might be responsible for a phase-variable absorption feature if temperature, elemental abundance or magnetic field vary strongly on the surface. The line energies observed in SGR 0418+5729 (~ 1 –5 keV) rule out transitions in magnetized H and He, which occur below ~ 1 keV (refs 24, 25). On the other hand, the absorption spectra of heavier elements are much more complex (see, for example, ref. 26 for C, O and Ne) and some lines

¹Istituto Universitario di Studi Superiori, piazza della Vittoria 15, I-27100 Pavia, Italy. ²Istituto di Astrofisica Spaziale e Fisica Cosmica Milano, INAF, via E. Bassini 15, I-20133 Milano, Italy. ³Istituto Nazionale di Fisica Nucleare, Sezione di Pavia, via A. Bassi 6, I-27100 Pavia, Italy. ⁴Dipartimento di Fisica e Astronomia, Università di Padova, via F. Marzolo 8, I-35131 Padova, Italy. ⁵Mullard Space Science Laboratory, University College London, Holmbury St Mary, Dorking, Surrey RH5 6NT, UK. ⁶AIM CEA/Irfu/Service d'Astrophysique, Orme des Merisiers, F-91191 Gif-sur-Yvette, France. ⁷Osservatorio Astronomico di Roma, INAF, via Frascati 33, I-00040 Monteporzio Catone, Italy. ⁸Institut de Ciències de l'Espai (IEEC-CSIC), Campus UAB, Torre C5, 2a planta, E-08193 Barcelona, Spain.

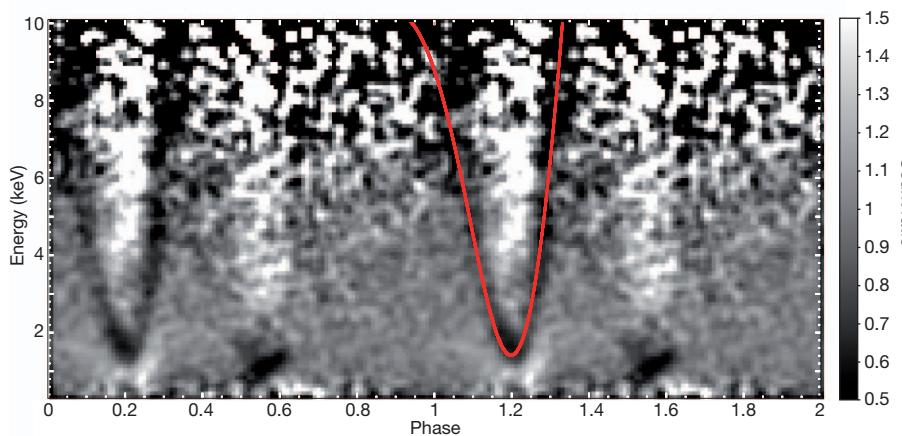


Figure 1 | Phase-dependent spectral feature in the EPIC data of SGR 0418+5729. Normalized energy versus phase image obtained by binning the EPIC source counts into 100 phase bins and 100-eV-wide energy channels and dividing these values first by the average number of counts in the same energy bin (corresponding to the phase-averaged energy spectrum) and then by the relative 0.3–10 keV count rate in the same phase interval (corresponding to the pulse profile normalized to the average count rate). The red line shows (for only one of the two displayed cycles) the results of a simple proton cyclotron model consisting of a baryon-loaded plasma loop emerging from the surface of a magnetar and intercepting the X-ray radiation from a small hotspot (see Supplementary Fig. 7 and Supplementary Table 1).

could occur at high enough energies. However, to explain the phase resolved spectra of SGR 0418+5729, the physical conditions of a heavy-element atmosphere are forced to vary in such a way that a single transition should dominate the opacity at each of the phases where the absorption line is detected.

A more straightforward explanation for the line variability can instead be given if the feature is due to cyclotron resonant scattering.

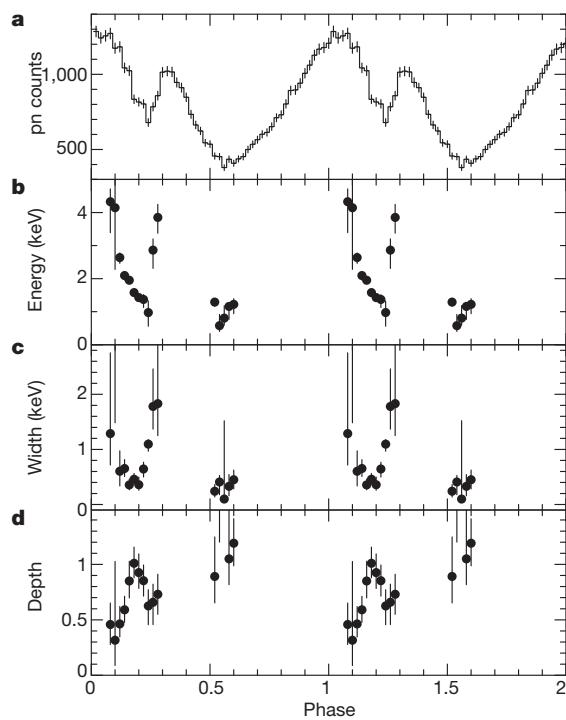


Figure 2 | Results of the phase-resolved spectroscopy of SGR 0418+5729. **a**, Pulse profile obtained by folding the 0.3–10 keV EPIC positive-negative junction (pn) detector light curve at the neutron star spin period $P = 9.07838827$ s. The data points are the number of counts in each phase-dependent spectrum. **b–d**, Line energy (E_c ; **b**), width (W ; **c**) and depth (D ; **d**) of the cyclotron feature as a function of the spin phase. The model consists of a blackbody plus a power law and an absorption line, modified for the interstellar absorption (see Supplementary Information). For the line we used the cyclotron absorption model from ref. 15: $F(E) = \exp\left(-D \frac{(WE/E_c)^2}{(E-E_c)^2 + W^2}\right)$. The interstellar absorption, temperature, photon index and relative normalizations of the two components were fixed to the best-fit values of the phase-integrated spectrum: $N_H = (9.6 \pm 0.5) \times 10^{21} \text{ cm}^{-2}$, $kT = 913 \pm 8 \text{ eV}$, $\Gamma = 2.8 \pm 0.2$, $(R_{BB}/d)^2 = 0.81 \pm 0.03 \text{ km}^2 \text{ kpc}^{-2}$ and $K_{PL} = (1.5 \pm 0.2) \times 10^{-3} \text{ photons cm}^{-2} \text{ s}^{-1} \text{ keV}^{-1}$ at 1 keV. Vertical error bars, 1s.d.

The cyclotron energy (in keV) for a particle of charge e and mass m in magnetic field B (in gauss) is given by

$$E_B \approx \frac{11.6}{1+z} \left(\frac{m_e}{m}\right) \left(\frac{B}{10^{12}}\right)$$

where $(1+z)^{-1} = [1 - 2GM_{NS}/(Rc)^2]^{1/2}$ (which is ~ 0.8 at the star surface for typical neutron star mass and radius $M_{NS} = 1.4M_\odot$ and $R_{NS} = 12 \text{ km}$, respectively) accounts for the gravitational redshift at distance R from the neutron star centre, m_e is the mass of the electron, and c is the velocity of light. In this case, the phase variability of the feature energy would simply be due to the different fields experienced by the charged particles interacting with the photons directed towards us as the neutron star rotates.

If the absorbers and scatterers are electrons hovering near the star surface, the expected line energy is ~ 70 keV for the dipole field at the equator of SGR 0418+5729 ($B = 6 \times 10^{12} \text{ G}$); this line energy is more than 10 times higher than that observed. A possible way to explain this large discrepancy might be to assume that the electrons producing the line are located higher up in the magnetosphere in a dipolar geometry, where the magnetic field is smaller ($R \approx 3R_{NS}$ to have $E_B \approx 2$ keV). Moreover, such an electron population should also be nearly monoenergetic, or subrelativistic, in order to prevent Compton scattering from washing out the feature, which would require a mechanism to maintain slowly moving electrons confined in a small volume high in the magnetosphere.

If the particles responsible for the cyclotron scattering are protons, the energy range of the SGR 0418+5729 spectral feature requires a magnetic field $> 2 \times 10^{14} \text{ G}$ (it would be even larger for heavier ions). In the framework of the magnetar model, the unprecedented phase-variability of the line energy can be explained by the complex topology of magnetar magnetospheres, in which global and/or localized twists play an important part⁵. This is particularly true for SGR 0418+5729, which has a weak dipolar component, as testified by the small spin-down value, whereas a much stronger internal magnetic field has been advocated to explain its X-ray luminosity and burst activity^{10,11}. Furthermore, the presence of small-scale, strong, multipolar components of the surface field has been inferred by fitting its phase-averaged X-ray spectrum with models of magnetized neutron star atmospheres¹².

In this context, the observed line variability might be due to the presence of either strong magnetic field gradients along the surface or vertical structures (with a spatially dependent field) emerging from the surface. To work out how the dynamic magnetosphere of a magnetar should look, an analogy with the solar corona in the proximity of sunspots has been proposed (see, for example, ref. 27). In particular, localized, baryon-rich magnetic structures (in the form of rising flux tubes) or ‘prominences’ (produced by magnetic reconnection or the emergence of the internal field near a crustal fault) have been proposed to explain some of the observed properties of the giant flare emitted

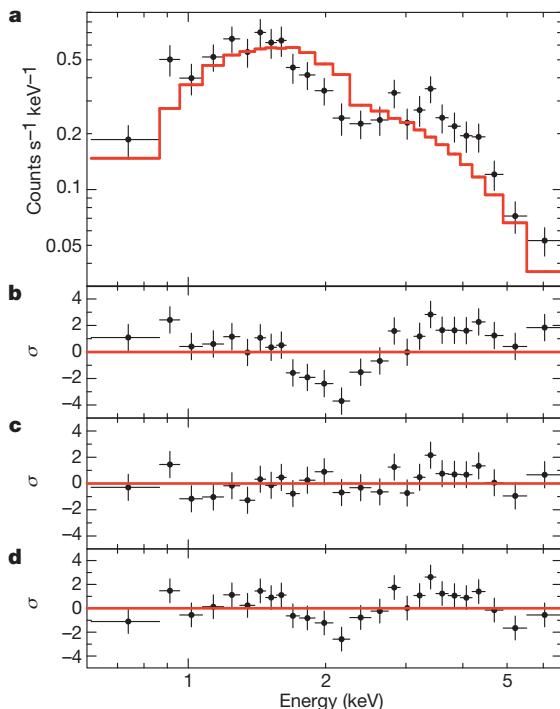


Figure 3 | Example of a phase-resolved EPIC pn spectrum and its residuals with respect to different models. **a**, Spectrum from the phase interval 0.15–0.17 (black dots) and best-fit model of the phase-averaged spectrum, rescaled with a free normalization factor (red line). **b**, Residuals with respect to this model ($\chi^2_{\nu} = 2.75$ for 25 d.f.); **c**, residuals after the addition of an absorption line (cyclabs model in XSPEC, with parameters as in Fig. 2; $\chi^2_{\nu} = 0.94$ for 22 d.f.); **d**, residuals with respect to an absorbed blackbody plus power-law model with free temperature, photon index and normalizations ($kT = 1.11 \pm 0.06$ keV and $\Gamma = 3.8 \pm 0.4$; $\chi^2_{\nu} = 1.75$ for 22 d.f.). This is one of the models (with the same number of free parameters) that we also explored to fit the phase-resolved spectra. In this case, we obtained fits of comparable quality to those with the line model at most phases, but worse fits in the phase interval 0.11–0.21. A joint fit to these five spectra gave an unacceptable χ^2_{ν} of 1.56 for 116 d.f., to be compared with χ^2_{ν} of 1.08 for the absorption line model (same number of d.f.). Horizontal error bars indicate the energy channel width; vertical error bars, 1 s.d.; residuals σ are in units of standard deviations.

in 2004 by SGR 1806–20^{28,29}. If a similar scenario, albeit on a reduced scale, occurred during the outburst of SGR 0418+5729, a spectral feature might arise as thermal photons from the hotspot (a small hot region on the neutron star surface, responsible for most of the X-ray emission, which could be itself related to the prominence) cross the plasma threading the magnetic loop. A proton density $\sim 10^{17}$ cm⁻³ is needed to produce a resonant scattering depth of order unity⁵. Protons, being heavy, do not rise much above the surface and move subrelativistically⁵, so resonant scattering in the prominence is likely to produce a narrow feature instead of an extended tail. As the star rotates, photons emitted in different directions pass through portions of the prominence with different magnetic field, density and size, giving rise to the observed variations of the line centroid and width. A simple quantitative model based on this picture is presented in Supplementary Information. Results, obtained with a geometry consistent with the constraints derived from the X-ray pulsed fraction of SGR 0418+5729, are in good agreement with the observed variations of the feature with phase (Fig. 1).

Received 8 March; accepted 11 June 2013.

- Mereghetti, S. The strongest cosmic magnets: soft gamma-ray repeaters and anomalous X-ray pulsars. *Astron. Astrophys. Rev.* **15**, 225–287 (2008).

- Rea, N. & Esposito, P. in *High-Energy Emission from Pulsars and their Systems* (eds Torres, D. F. & Rea, N.) 247–273 (Astrophysics and Space Science Proceedings, Springer, 2011).
- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars—I. Radiative mechanism for outbursts. *Mon. Not. R. Astron. Soc.* **275**, 255–300 (1995).
- Thompson, C. & Duncan, R. C. The soft gamma repeaters as very strongly magnetized neutron stars. II. Quiescent neutrino, X-ray, and Alfvén wave emission. *Astrophys. J.* **473**, 322–342 (1996).
- Thompson, C., Lyutikov, M. & Kulkarni, S. R. Electrodynamics of magnetars: implications for the persistent X-ray emission and spin-down of the soft gamma repeaters and anomalous X-ray pulsars. *Astrophys. J.* **574**, 332–355 (2002).
- Kouveliotou, C. et al. An X-ray pulsar with a superstrong magnetic field in the soft gamma-ray repeater SGR 1806–20. *Nature* **393**, 235–237 (1998).
- Thompson, C. & Duncan, R. C. The giant flare of 1998 August 27 from SGR 1900+14. II. Radiative mechanism and physical constraints on the source. *Astrophys. J.* **561**, 980–1005 (2001).
- Vietri, M., Stella, L. & Israel, G. L. SGR 1806–20: evidence for a superstrong magnetic field from quasi-periodic oscillations. *Astrophys. J.* **661**, 1089–1093 (2007).
- Rea, N. et al. The outburst decay of the low magnetic field magnetar SGR 0418+5729. *Astrophys. J.* **770**, 65 (2013).
- Rea, N. et al. A low-magnetic-field soft gamma repeater. *Science* **330**, 944 (2010).
- Turolla, R., Zane, S., Pons, J. A., Esposito, P. & Rea, N. Is SGR 0418+5729 indeed a waning magnetar? *Astrophys. J.* **740**, 105 (2011).
- Güver, T., Göğüş, E. & Özel, F. A magnetar strength surface magnetic field for the slowly spinning down SGR 0418+5729. *Mon. Not. R. Astron. Soc.* **418**, 2773–2778 (2011).
- van der Horst, A. J. et al. Discovery of a new soft gamma repeater: SGR J0418+5729. *Astrophys. J.* **711**, L1–L6 (2010).
- Esposito, P. et al. Early X-ray and optical observations of the soft gamma-ray repeater SGR0418+5729. *Mon. Not. R. Astron. Soc.* **405**, 1787–1795 (2010).
- Makishima, K. et al. Observations of the peculiar hard X-ray transient X0331+53 (V0332+53). *Publ. Astron. Soc. Jpn* **42**, 295–315 (1990).
- Truemper, J. et al. Evidence for strong cyclotron line emission in the hard X-ray spectrum of Hercules X-1. *Astrophys. J.* **219**, L105–L110 (1978).
- Heindl, W. A. et al. in *X-ray Timing 2003: Rossi and Beyond* (eds Kaaret, P., Lamb, F. K. & Swank, J. H.) 323–330 (AIP Conf. Ser. Vol. 714, American Institute of Physics, 2004).
- Turolla, R. in *Neutron Stars and Pulsars* (ed. Becker, W.) 141–163 (Astrophysics and Space Science Proceedings Vol. 357, Springer, 2009).
- Haberl, F. et al. Evidence for precession of the isolated neutron star RX J0720.4–3125. *Astron. Astrophys.* **451**, L17–L21 (2006).
- van Kerkwijk, M. H. & Kaplan, D. L. Isolated neutron stars: magnetic fields, distances, and spectra. *Astrophys. Space Sci.* **308**, 191–201 (2007).
- Bignami, G. F., Caraveo, P. A., De Luca, A. & Mereghetti, S. The magnetic field of an isolated neutron star from X-ray cyclotron absorption lines. *Nature* **423**, 725–727 (2003).
- Gotthelf, E. V., Halpern, J. P. & Alföldi, J. The spin-down of PSR J0821–4300 and PSR J1210–5226: confirmation of central compact objects as anti-magnetars. *Astrophys. J.* **765**, 58 (2013).
- Kargaltsev, O., Durant, M., Misanovic, Z. & Pavlov, G. G. Absorption features in the X-ray spectrum of an ordinary radio pulsar. *Science* **337**, 946–949 (2012).
- Potekhin, A. Y. Hydrogen atom moving across a strong magnetic field: analytical approximations. *J. Phys. At. Mol. Opt. Phys.* **31**, 49–63 (1998).
- Medin, Z., Lai, D. & Potekhin, A. Y. Radiative transitions of the helium atom in highly magnetized neutron star atmospheres. *Mon. Not. R. Astron. Soc.* **383**, 161–172 (2008).
- Mori, K. & Ho, W. C. G. Modelling mid-Z element atmospheres for strongly magnetized neutron stars. *Mon. Not. R. Astron. Soc.* **377**, 905–919 (2007).
- Beloborodov, A. M. & Thompson, C. Corona of magnetars. *Astrophys. J.* **657**, 967–993 (2007).
- Gelfand, J. D. et al. A rebrightening of the radio nebula associated with the 2004 December 27 giant flare from SGR 1806–20. *Astrophys. J.* **634**, L89–L92 (2005).
- Masada, Y., Nagataki, S., Shibata, K. & Terasawa, T. Solar-type magnetic reconnection model for magnetar giant flares. *Publ. Astron. Soc. Jpn.* **62**, 1093–1102 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank G. Goggi and C. Paizis for discussions. This work is based on data and software provided by the ESA XMM-Newton Science Archive (XSA) and the NASA/GSFC High Energy Astrophysics Science Archive Research Center (HEASARC). We acknowledge partial funding from INAF through a PRIN 2010 grant and ASI through contract I/032/10/0.

Author Contributions A.T., P.E. and S.M. processed and analysed the data, with contributions by F.G. and D.G.; theoretical interpretation was provided by R.T., L.N. and A.T., with contributions by S.Z. and L.S.; A.T., P.E., S.M. and R.T. composed the text, using inputs from the other co-authors; all authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T. (andrea.tiengo@iusspavia.it).

Deterministic quantum teleportation of photonic quantum bits by a hybrid technique

Shuntaro Takeda¹, Takahiro Mizuta¹, Maria Fuwa¹, Peter van Loock² & Akira Furusawa¹

Quantum teleportation¹ allows for the transfer of arbitrary unknown quantum states from a sender to a spatially distant receiver, provided that the two parties share an entangled state and can communicate classically. It is the essence of many sophisticated protocols for quantum communication and computation^{2–5}. Photons are an optimal choice for carrying information in the form of ‘flying qubits’, but the teleportation of photonic quantum bits^{6–11} (qubits) has been limited by experimental inefficiencies and restrictions. Main disadvantages include the fundamentally probabilistic nature of linear-optics Bell measurements¹², as well as the need either to destroy the teleported qubit or attenuate the input qubit when the detectors do not resolve photon numbers¹³. Here we experimentally realize fully deterministic quantum teleportation of photonic qubits without post-selection. The key step is to make use of a hybrid technique involving continuous-variable teleportation^{14–16} of a discrete-variable, photonic qubit. When the receiver’s feedforward gain is optimally tuned, the continuous-variable teleporter acts as a pure loss channel^{17,18}, and the input dual-rail-encoded qubit, based on a single photon, represents a quantum error detection code against photon loss¹⁹ and hence remains completely intact for most teleportation events. This allows for a faithful qubit transfer even with imperfect continuous-variable entangled states: for four qubits the overall transfer fidelities range from 0.79 to 0.82 and all of them exceed the classical limit of teleportation. Furthermore, even for a relatively low level of the entanglement, qubits are teleported much more efficiently than in previous experiments, albeit post-selectively (taking into account only the qubit subspaces), and with a fidelity comparable to the previously reported values.

Since originally being proposed¹, the concept of quantum teleportation has attracted a lot of attention and has become one of the central elements of advanced and practical realizations of quantum information protocols. It is essential for long-distance quantum communication by means of quantum repeaters², and it has been shown to be a useful tool for realizing universal quantum logic gates in a measurement-based fashion³. Many proposals and models for quantum computation rely on quantum teleportation, such as the efficient linear-optics quantum computing scheme of ref. 4 and the ‘one-way’ quantum computer using cluster states⁵.

Although much progress has been made in demonstrating quantum teleportation of photonic qubits^{6–11}, most of these schemes share one fundamental restriction: an unambiguous two-qubit Bell-state measurement (BSM), as is needed to teleport a qubit using two-qubit entanglement, is always probabilistic when linear optics is used, even if photon-number-resolving detectors are available^{12,13}. There are two experiments avoiding this constraint, but in these either a qubit can no longer be teleported when it is delivered independently from an external resource⁷ or an extra nonlinear element leads to extremely low measurement efficiencies, of the order of 10^{-10} (ref. 8). A further experimental limitation, rendering these schemes fairly inefficient, is the probabilistic nature of the entangled resource states¹³. Efficient, near-deterministic quantum teleportation, however, is of great benefit in quantum communication,

where it can be used to reduce the storage times of quantum memories in a quantum repeater, and it is a necessity in teleportation-based quantum computation. An additional drawback of the previous experiments, owing to the lack of photon-number-resolving detectors, was the need either to destroy the teleported qubit²⁰ or to attenuate the input qubit¹⁰, thus further decreasing the success rate of teleportation.

We overcome all the above limitations by taking a different approach: continuous-variable quantum teleportation of a photonic qubit. The strength of continuous-variable teleportation lies in the on-demand availability of the quadrature-entangled states and the completeness of a BSM in the quadrature bases using linear optics and homodyne detections¹⁵. So far, these tools have been used to unconditionally teleport continuous-variable quantum states such as coherent states^{16,21}. However, it has not yet been possible to apply them to qubits^{18,22}, because typical pulsed-laser-based qubits (such as those in refs 6–11) have a broad frequency bandwidth that is incompatible with the original continuous-wave-based continuous-variable teleporter, which works only on narrow sidebands^{16,21}. We overcome this incompatibility by using very recent, advanced technology: a broadband continuous-variable teleporter²³ and a narrowband time-bin qubit compatible with that teleporter²⁴. Importantly, this qubit uses two temporal modes to represent a dual-rail-encoded qubit¹³

$$|\psi\rangle = \alpha|0, 1\rangle + \beta|1, 0\rangle$$

where $|0, 1\rangle$ and $|1, 0\rangle$ refer to the temporal modes of the photon (expressed in the two-mode photon-number basis, with $|\alpha|^2 + |\beta|^2 = 1$). Therefore, teleportation of both modes of the qubit is accomplished by means of a single continuous-variable teleporter acting subsequently on the temporal modes of the time-bin qubits (Fig. 1).

Remarkably, the main weakness of continuous-variable teleportation, namely the intrinsic imperfection of the finitely squeezed, entangled states, can be circumvented to a great extent in the present ‘hybrid’ setting when the input to the continuous-variable teleporter is a dual-rail qubit. The entangled state of the teleporter is a two-mode squeezed, quadrature-entangled state, $\sqrt{1 - g_{\text{opt}}^2} \sum_{n=0}^{\infty} g_{\text{opt}}^n |n, n\rangle$, here written in the number basis with $g_{\text{opt}} = \tanh(r)$, where r is the squeezing parameter. Because infinite squeezing ($r \rightarrow \infty$) requires infinite energy, maximally entangled states are physically unattainable; thus, the teleportation fidelity is generally limited by r . Following the standard continuous-variable quantum teleportation protocol with unit gain for the receiver’s feedforward displacement¹⁵ yields a largely distorted output qubit with additional thermal photons.

In contrast, non-unit gain conditions are useful in some cases^{25,26}. In particular, a single-mode continuous-variable teleporter with gain g_{opt} creates no additional photons, because it is equivalent to a pure attenuation channel from which an intensity fraction of $1 - g_{\text{opt}}^2$ is lost to the environment^{17,18}. Moreover, the dual-rail-encoded qubit represents a quantum error detection code against such amplitude damping, where either a photon-loss error occurs, erasing the qubit, or a symmetric damping leaves the input qubit state completely intact¹⁹. These two facts together mean that the dual-rail continuous-variable teleporter at

¹Department of Applied Physics, School of Engineering, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. ²Institute of Physics, Johannes-Gutenberg Universität Mainz, Staudingerweg 7, 55128 Mainz, Germany.

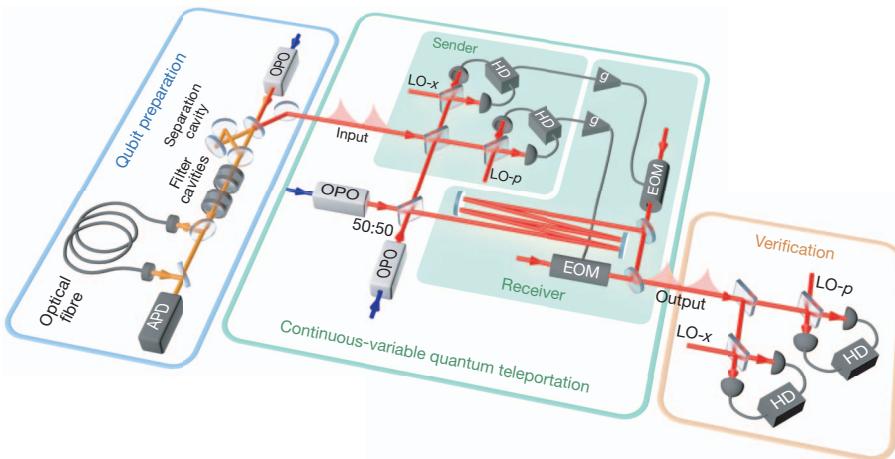


Figure 1 | Experimental set-up. A time-bin qubit is heralded by detecting one half of an entangled photon pair produced by an optical parametric oscillator (OPO). The continuous-variable teleporter (g , feedforward gain) always transfers this qubit, albeit with non-unit fidelity. The teleported qubit is finally

optimal gain, g_{opt} , transforms the initial qubit state as follows:

$$|\psi\rangle\langle\psi| \rightarrow g_{\text{opt}}^2 |\psi\rangle\langle\psi| + (1 - g_{\text{opt}}^2) |0, 0\rangle\langle 0, 0| \quad (1)$$

Most importantly, no additional photons are created and the quantum information encoded in $|\psi\rangle$ remains undisturbed regardless of the squeezing level. The only effect of the teleporter is the extra two-mode vacuum term, whose fraction becomes arbitrarily small for sufficiently large squeezing, $g_{\text{opt}} \rightarrow 1$. This technique allows us to teleport arbitrary qubit states more faithfully by suppressing additional photons, thereby realizing unconditional teleportation with a moderate level of squeezing. Equation (1) also shows that a fidelity of unity is obtainable for any non-zero squeezing level, $g_{\text{opt}} > 0$, provided that the signal qubit subspace is post-selected, that is, the non-occurrence of a photon-loss error is detected with a probability approaching zero for $g_{\text{opt}} \rightarrow 0$. We note that the remaining vacuum contribution could be made arbitrarily small without post-selection of the final states, by instead immediately discarding all quadrature results of the BSM that are too far from the phase-space origin^{22,27}.

To demonstrate successful qubit quantum teleportation, we prepare four distinct qubit states: $|0, 1\rangle, |1, 0\rangle, |\psi_1\rangle \equiv (|0, 1\rangle - i|1, 0\rangle)/\sqrt{2}$ and $|\psi_2\rangle \equiv (2|0, 1\rangle - |1, 0\rangle)/\sqrt{5}$. This set, including even and uneven superpositions of $|0, 1\rangle$ and $|1, 0\rangle$ with both real and imaginary phases, represents a fair sample of qubit states on the Bloch sphere. In theory, our teleporter acts on any qubit state in the same way (Supplementary Discussion).

The experimental density matrix of the input state, $|\psi_1\rangle$, is shown in Fig. 2a. This input state is not a pure qubit state, but rather a mixed state with a $25 \pm 1\%$ vacuum contribution, a $69 \pm 1\%$ qubit contribution and a $6 \pm 1\%$ multiphoton contribution. Because the continuous-variable teleporter transfers input states of arbitrary dimension, all of these components are teleported and constitute the final, mixed output state. We note that in our first analysis we do not discard any of these contributions from the input or the output states, thus ensuring that none of the quantum states that enter or leave our teleporter is pre-selected or post-selected, respectively.

First we present the output state of unit-gain teleportation with $r = 1.01 \pm 0.03$ (Fig. 2b). All the matrix elements obtained are in good agreement with theory: the qubit contribution decreases, whereas the contribution of the multiphoton terms grows owing to the finite squeezing. The off-diagonal elements of the qubit ($|0, 1\rangle\langle 1, 0|, |1, 0\rangle\langle 0, 1|$) retain the original phase information of the input superposition between $|0, 1\rangle$ and $|1, 0\rangle$, demonstrating that the non-classical feature of the qubits is preserved during the teleportation process. These off-diagonal elements, however, decay a little more rapidly than do the diagonal

characterized by single or dual homodyne measurement to verify the success of teleportation. See Methods Summary for details. APD, avalanche photodiode; EOM, electro-optic modulator; HD, homodyne detector; LO- x and LO- p , local oscillators to measure x and p quadratures, respectively.

elements ($|0, 1\rangle\langle 0, 1|, |1, 0\rangle\langle 1, 0|$), illustrating that the quantum superposition of the qubit is the feature most susceptible to error in an experimental situation.

Next we turned down the gain, g , and observed the new output state. Figure 2c shows the output state at $g = 0.79$ (close to $g_{\text{opt}} = 0.77$). Compared with Fig. 2b, from Fig. 2c it can be seen that the qubit components are almost undisturbed, but that the vacuum grows and the occurrence of extra multiphoton components is suppressed. Thus, here the input-output relation is similar to the pure-attenuation model with a loss fraction of $1 - g_{\text{opt}}^2 = 0.41$. The bar graph in Fig. 3 shows the g dependence of the qubit and multiphoton components in the output state, clearly demonstrating that gain tuning reduces the creation of additional photons in continuous-variable teleportation.

The performance of the teleporter can be assessed by means of the fidelity. In our deterministic scheme, we must take into account the vacuum and multiphoton contributions, which was not the case in previous non-deterministic teleportation experiments using post-selection. The overall fidelity between the input state, $\hat{\rho}_{\text{in}}$, and the output state, $\hat{\rho}_{\text{out}}$, is²⁸

$$F_{\text{state}} = \left[\text{Tr} \left(\sqrt{\sqrt{\hat{\rho}_{\text{in}}} \hat{\rho}_{\text{out}} \sqrt{\hat{\rho}_{\text{in}}}} \right) \right]^2$$

When $\hat{\rho}_{\text{in}}$ has a qubit fraction of η , the classical bound on F_{state} is $F_{\text{thr}} \equiv 1 - \eta/3$, which is the best fidelity achievable without entanglement (Supplementary Discussion). Therefore, $F_{\text{state}} > F_{\text{thr}}$ is a success criterion for unconditional quantum teleportation. Alternatively, we may also assess our teleporter by calculating the fidelity after post-selecting the qubit components alone: $F_{\text{qubit}} = \langle \psi | \hat{\rho}_{\text{out}}^{\text{qubit}} | \psi \rangle$, where $|\psi\rangle$ is the ideal qubit state and $\hat{\rho}_{\text{out}}^{\text{qubit}}$ is obtained by extracting from the output density matrix the qubit subspace spanned by $\{|0, 1\rangle, |1, 0\rangle\}$ and then renormalizing it. We note that $F_{\text{qubit}} > 2/3$ is the success criterion of post-selective teleportation with a pure input qubit and a mixed output qubit²⁹.

As shown in Fig. 3, the g dependence of these two fidelities is in good agreement with the theoretical predictions. The maximal fidelities are obtained at $g = 0.79$. Most importantly, here, we satisfy not only the usual qubit-subspace teleportation criterion, $F_{\text{qubit}} = 0.875 \pm 0.015 > 2/3$, but also the fully non-post-selected, Fock-space criterion, $F_{\text{state}} = 0.817 \pm 0.012 > F_{\text{thr}} = 0.769 \pm 0.004$, thus demonstrating deterministic, unconditional quantum teleportation of a photonic qubit. As well as for the input qubit, $|\psi_1\rangle$, the Fock-space criterion is also fulfilled for the other three qubit states, $|0, 1\rangle, |1, 0\rangle$ and $|\psi_2\rangle$, with the same experimental r and g values, for which states F_{state} values of $0.800 \pm 0.006, 0.789 \pm 0.006$ and 0.796 ± 0.011 were respectively observed (theoretically, F_{state} and F_{qubit} are independent of the

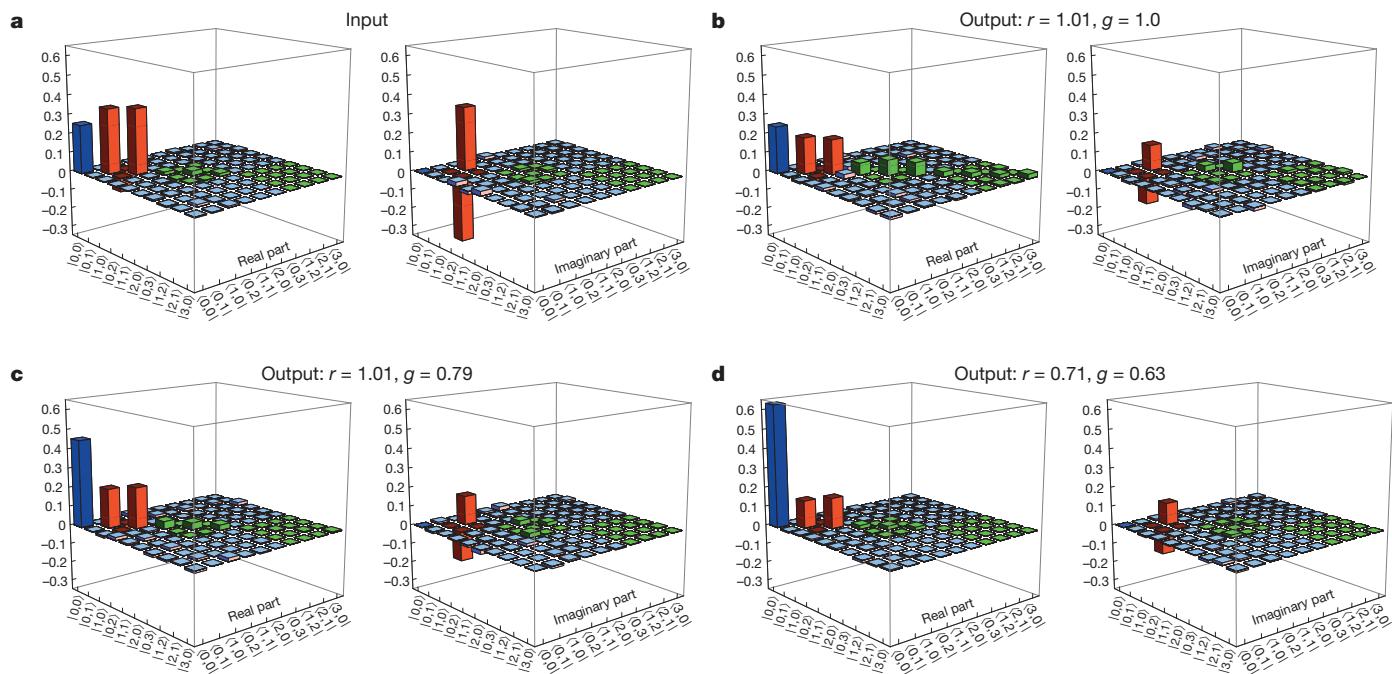


Figure 2 | Experimental density matrices. By means of homodyne tomography, two-mode density matrices are reconstructed both for the input and the output states in photon-number bases²⁴: $\hat{p} = \sum_{k,l,m,n=0}^{\infty} \rho_{klmn} |k, l\rangle \langle m, n|$. The bars show the real or imaginary parts

qubit; see Supplementary Discussion and Supplementary Data). We note that, although the pure-attenuation model predicts $F_{\text{qubit}} = 1$ and a complete suppression of multiphoton terms at gain g_{opt} , our results slightly deviate from that situation owing to experimental imperfections, such as extra loss and phase fluctuations of the squeezing.

Finally, Fig. 2d shows the output state for the lower squeezing level, $r = 0.71 \pm 0.02$, and $g = 0.63$ ($g_{\text{opt}} = 0.61$). Here, although the vacuum component becomes more dominant, the qubit components retain almost the same form as in Fig. 2c. Under these circumstances, the success of teleportation is only post-selective ($F_{\text{qubit}} = 0.879 \pm 0.015 > 2/3$, $F_{\text{state}} < F_{\text{thr}}$), owing to the insufficient squeezing resource. However, the overall success probability for transferring the qubits ($43 \pm 3\%$, the ratio of

of each matrix element ρ_{klmn} . Blue, red and green bars correspond to the vacuum, qubit and multiphoton components, respectively. **a**, Input state, $|\psi_1\rangle$. **b–d**, Output states for different values of r and g .

the input and output qubit components) is still much higher than in previous experiments ($\ll 1\%$), even for this relatively low squeezing level. This shows the great advantage of our hybrid approach over the standard approaches.

In conclusion, we experimentally realized unconditional quantum teleportation of four distinct photonic qubit states, exceeding the fidelity limits of classical teleportation in a deterministic fashion. In our scheme, once the input qubit states are prepared, there is no need to preprocess or post-select them, and the teleported states freely emerge at the output of our teleporter.

METHODS SUMMARY

Our experimental set-up is shown in Fig. 1. The time-bin qubit is conditionally created at a rate of $\sim 5,000 \text{ s}^{-1}$ using a continuous-wave laser²⁴ (wavelength, 860 nm), by extending the technique of ref. 30. Each time bin has a frequency bandwidth of 6.2 MHz around the laser frequency. Our continuous-variable teleporter²³ operates continuously with a bandwidth of 12 MHz around the laser frequency, which is sufficiently wide to cover the qubit bandwidth—ultimately enabling us to teleport qubits in a deterministic fashion. In our teleporter, two single-mode squeezed states (each with an ideal, pure squeezing parameter r) from two optical parametric oscillators are suitably mixed at a 50:50 beam splitter to generate the quadrature-entangled beams. This entanglement source is permanently available with no need for any probabilistic heralding mechanism. At the sending station of the teleporter, the input qubit is first combined with one half of the entangled beams at a 50:50 beam splitter. A complete continuous-variable BSM is then performed by measuring the two output modes of the beam splitter through two homodyne detections of two orthogonal quadratures. These homodyne signals are classically communicated to the receiving station, where they are multiplied by a gain factor (g) and fed forwards by means of a displacement operation on the other half of the entangled beams. Time synchronization of this final displacement is achieved by introducing an optical delay to the corresponding entangled beam. Finally, the output state is characterized by single or dual homodyne measurement²⁴. For every state, 100,000 sets of quadrature values are recorded and the corresponding two-mode density matrix is reconstructed using the maximum-likelihood technique without compensating finite measurement efficiencies.

Received 19 March; accepted 7 June 2013.

1. Bennett, C. H. et al. Teleporting an unknown quantum state via dual classical and Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).

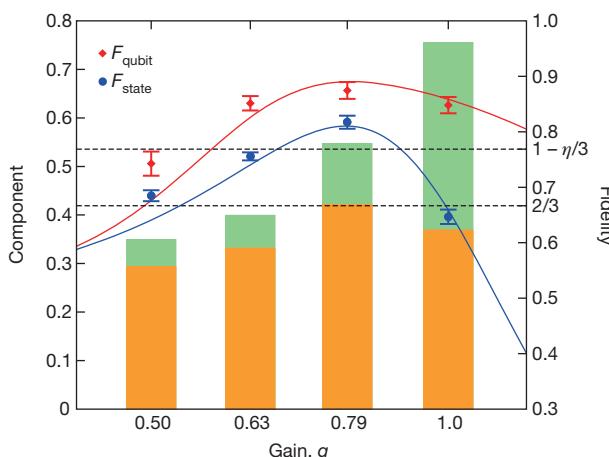


Figure 3 | Experimental results of teleportation including gain tuning. The horizontal axis, showing g , uses a logarithmic scale. Orange and green bars respectively represent qubit and multiphoton components of the teleported states (the left-hand vertical axis). Red diamonds and blue circles with error bars (1 s.d.) correspond to F_{qubit} and F_{state} , respectively (the right-hand vertical axis). Theoretical fidelity curves (Supplementary Information) are also plotted, in the same colours. All observed F_{qubit} values significantly exceed the classical limit of $2/3$. For $g = 0.79$, $F_{\text{state}} > 1 - \eta/3$ and, thus, unconditional teleportation is demonstrated.

2. Briegel, H.-J., Dür, W., Cirac, J. I. & Zoller, P. Quantum repeaters: the role of imperfect local operations in quantum communication. *Phys. Rev. Lett.* **81**, 5932–5935 (1998).
3. Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
4. Knill, E., Laflamme, R. & Milburn, G. J. A scheme for efficient quantum computation with linear optics. *Nature* **409**, 46–52 (2001).
5. Raussendorf, R. & Briegel, H. J. A one-way quantum computer. *Phys. Rev. Lett.* **86**, 5188–5191 (2001).
6. Bouwmeester, D. *et al.* Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
7. Boschi, D., Branca, S., De Martini, F., Hardy, L. & Popescu, S. Experimental realization of teleporting an unknown pure quantum state via dual classical Einstein-Podolsky-Rosen channels. *Phys. Rev. Lett.* **80**, 1121–1125 (1998).
8. Kim, Y.-H., Kulik, S. P. & Shih, Y. Quantum teleportation of polarization state with a complete Bell state measurement. *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
9. Marcikic, I., de Riedmatten, H., Tittel, W., Zbinden, H. & Gisin, N. Long-distance teleportation of qubits at telecommunication wavelengths. *Nature* **421**, 509–513 (2003).
10. Pan, J.-W., Gasparoni, S., Aspelmeyer, M., Jennewein, T. & Zeilinger, A. Experimental realization of freely propagating teleported qubits. *Nature* **421**, 721–725 (2003).
11. Ma, X.-S. *et al.* Quantum teleportation over 143 kilometres using active feed-forward. *Nature* **489**, 269–273 (2012).
12. Lütkenhaus, N., Calsamiglia, J. & Suominen, K.-A. Bell measurements for teleportation. *Phys. Rev. A* **59**, 3295–3300 (1999).
13. Pan, J.-W. *et al.* Multiphoton entanglement and interferometry. *Rev. Mod. Phys.* **84**, 777–838 (2012).
14. Vaidman, L. Teleportation of quantum states. *Phys. Rev. A* **49**, 1473–1476 (1994).
15. Braunstein, S. L. & Kimble, H. J. Teleportation of continuous quantum variables. *Phys. Rev. Lett.* **80**, 869–872 (1998).
16. Furusawa, A. *et al.* Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
17. Hofmann, H. F., Ide, T., Kobayashi, T. & Furusawa, A. Information losses in continuous-variable quantum teleportation. *Phys. Rev. A* **64**, 040301(R) (2001).
18. Polkinghorne, R. E. S. & Ralph, T. C. Continuous variable entanglement swapping. *Phys. Rev. Lett.* **83**, 2095–2099 (1999).
19. Nielsen, M. A. & Chuang, I. L. *Quantum Computation and Quantum Information* 380–386 (Cambridge Univ. Press, 2000).
20. Braunstein, S. L. & Kimble, H. J. *A posteriori* teleportation. *Nature* **394**, 840–841 (1998).
21. Yonezawa, H., Braunstein, S. L. & Furusawa, A. Experimental demonstration of quantum teleportation of broadband squeezing. *Phys. Rev. Lett.* **99**, 110503 (2007).
22. Ide, T., Hofmann, H. F., Kobayashi, T. & Furusawa, A. Continuous-variable teleportation of single-photon states. *Phys. Rev. A* **65**, 012313 (2001).
23. Lee, N. *et al.* Teleportation of nonclassical wave packets of light. *Science* **332**, 330–333 (2011).
24. Takeda, S. *et al.* Generation and eight-port homodyne characterization of time-bin qubits for continuous-variable quantum information processing. *Phys. Rev. A* **87**, 043803 (2013).
25. Bowen, W. P. *et al.* Experimental investigation of continuous-variable quantum teleportation. *Phys. Rev. A* **67**, 032302 (2003).
26. Jia, X. *et al.* Experimental demonstration of unconditional entanglement swapping for continuous variables. *Phys. Rev. Lett.* **93**, 250503 (2004).
27. Mišta, L. Jr, Filip, R. & Furusawa, A. Continuous-variable teleportation of a negative Wigner function. *Phys. Rev. A* **82**, 012322 (2010).
28. Jozsa, R. Fidelity for mixed quantum states. *J. Mod. Opt.* **41**, 2315–2323 (1994).
29. Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
30. Zavatta, A., D’Angelo, M., Parigi, V. & Bellini, M. Remote preparation of arbitrary time-encoded single-photon ebits. *Phys. Rev. Lett.* **96**, 020502 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was partly supported by PDIS, GIA, G-COE, APSA and FIRST, commissioned by MEXT (Japan); by the SCOPE programme commissioned by MIC (Japan); and by ASCR-JSPS. S.T. and M.F. acknowledge financial support from ALPS. We thank L. Mišta Jr, H. Yonezawa and J. Kimble for comments.

Author Contributions A.F. planned and supervised the project. P.v.L. and S.T. theoretically defined the scientific goals. S.T. and T.M. designed and performed the experiment, and acquired the data. S.T. and M.F. developed the electronic devices. S.T., T.M. and M.F. analysed the data. S.T. and P.v.L. wrote the manuscript with assistance from all other co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.F. (akiraf@ap.tu-tokyo.ac.jp).

Deterministic quantum teleportation with feed-forward in a solid state system

L. Steffen¹, Y. Salathe¹, M. Oppliger¹, P. Kurpiers¹, M. Baur^{1†}, C. Lang¹, C. Eichler¹, G. Puebla-Hellmann¹, A. Fedorov^{1†} & A. Wallraff¹

Engineered macroscopic quantum systems based on superconducting electronic circuits are attractive for experimentally exploring diverse questions in quantum information science^{1–3}. At the current state of the art, quantum bits (qubits) are fabricated, initialized, controlled, read out and coupled to each other in simple circuits. This enables the realization of basic logic gates⁴, the creation of complex entangled states^{5,6} and the demonstration of algorithms⁷ or error correction⁸. Using different variants of low-noise parametric amplifiers⁹, dispersive quantum non-demolition single-shot readout of single-qubit states with high fidelity has enabled continuous¹⁰ and discrete¹¹ feedback control of single qubits. Here we realize full deterministic quantum teleportation with feed-forward in a chip-based superconducting circuit architecture^{12–14}. We use a set of two parametric amplifiers for both joint two-qubit and individual qubit single-shot readout, combined with flexible real-time digital electronics. Our device uses a crossed quantum bus technology that allows us to create complex networks with arbitrary connecting topology in a planar architecture. The deterministic teleportation process succeeds with order unit probability for any input state, as we prepare maximally entangled two-qubit states as a resource and distinguish all Bell states in a single two-qubit measurement with high efficiency and high fidelity. We teleport quantum states between two macroscopic systems separated by 6 mm at a rate of 10^4 s⁻¹, exceeding other reported implementations. The low transmission loss of superconducting waveguides is likely to enable the range of this and other schemes to be extended to significantly larger distances, enabling tests of non-locality and the realization of elements for quantum communication at microwave frequencies. The demonstrated feed-forward may also find application in error correction schemes.

Transferring the state of an information carrier from a sender to a receiver is an essential primitive in both classical and quantum communication and information processing. In a quantum mechanical context, teleportation describes the process of transferring an unknown quantum state between two parties at two different physical locations without transferring the physical carrier of information itself. Instead, teleportation makes use of the non-local correlations provided by an entangled pair shared between the sender and the receiver and the exchange of classical information¹⁵. This concept is central to extending the range of quantum communication using quantum repeaters¹⁶ and can also be used to implement logic gates for universal quantum computation¹⁷.

In the original teleportation protocol¹⁵, the unknown state $|\psi_{in}\rangle$ of qubit Q1 in possession of the sender is transferred to the receiver's qubit, Q3 (Fig. 1a). To enable this task, sender and receiver prepare in advance a maximally entangled (Bell) state between an ancillary qubit Q2, which remains with the sender, and Q3, which is located at the receiver. Then the sender performs a measurement of Q1 and Q2 in the Bell basis, which projects the two qubits in the sender's possession onto one of the four possible Bell states $|\Phi^\pm\rangle = (|00\rangle \pm |11\rangle)/\sqrt{2}$ and $|\Psi^\pm\rangle = (|01\rangle \pm |10\rangle)/\sqrt{2}$. As a consequence the receiver's qubit Q3 is projected instantaneously and without ever having interacted with the sender's qubit Q1 onto a state $|\psi_{out}\rangle = \{\mathbb{1}, X, Z, \tilde{Y}\} |\psi_{in}\rangle$ (where

$X = \hat{\sigma}_x$, $Z = \hat{\sigma}_z$ and $\tilde{Y} = i\hat{\sigma}_y$ are the Pauli matrices), which differs from the input state $|\psi_{in}\rangle$ only by a single-qubit rotation, depending on the four possible measurement results. To always recover the original state $|\psi_{in}\rangle$ the receiver has to rotate the output state of Q3 conditioned on the outcome of the Bell measurement communicated to the receiver as two bits of information by means of a classical channel. This final step is frequently referred to as feed-forward, because the outcome of a measurement performed on one part of a system is used to control another part of the same quantum system. This is in contrast to acting back on the same quantum system in a feedback process.

The success of the teleportation protocol in every instance with unit fidelity is counterintuitive from a classical point of view. The receiver's qubit does not interact with any other qubit after $|\psi_{in}\rangle$ is prepared. The classical information sent by the sender is not sufficient to recreate $|\psi_{in}\rangle$ perfectly at the receiver. Indeed, assuming no entanglement between sender and receiver, one can replicate the sender's state at best with a fidelity of 2/3 (ref. 18) because only a fraction of information about $|\psi_{in}\rangle$ is obtained by a single projective measurement.

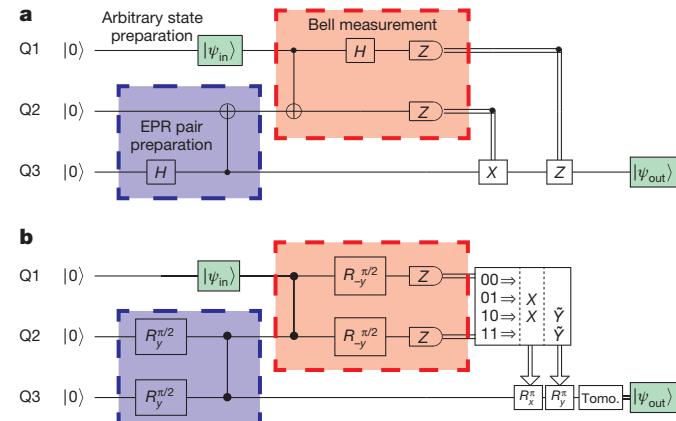


Figure 1 | Circuit diagram of quantum teleportation. **a**, The original protocol starts with the preparation of a Bell state between Q2 and Q3 (blue box), followed by the preparation of an arbitrary state $|\psi_{in}\rangle$ (green box) and a Bell-state measurement of Q1 and Q2 (red box). The classical information extracted by the measurement of Q1 and Q2 is transferred to the receiver to perform local gates conditioned on the measurement outcomes. After the protocol, Q3 is in state $|\psi_{out}\rangle$, which ideally is identical to $|\psi_{in}\rangle$ (also coloured green). Here, H is the Hadamard gate, and X and Z are Pauli matrices. The CNOT gate is represented by a vertical line between the control qubit (●) and the target qubit (⊕). **b**, The protocol implemented in our experiment uses controlled-PHASE gates indicated by vertical lines between the relevant qubits (●), and single-qubit rotations $R_{\pm y}^\theta$ of angle θ about the $\pm y$ axis. To finalize the teleportation we either post-select on any single one of the four measurement outcomes (00, 01, 10 and 11) acquired in a single shot, or we deterministically use all four outcomes, which we then may use to implement feed-forward. The feed-forward operators R_x^θ and R_y^θ are applied to Q3 conditioned on the four measurement outcomes according to the table presented in the box framed by a black line.

¹Department of Physics, ETH Zurich, CH-8093 Zurich, Switzerland. [†]Present address: ARC Centre for Engineered Quantum Systems, University of Queensland, Brisbane 4072, Australia.

In pioneering work the teleportation protocol was first implemented with single photons¹⁹ over laboratory-scale distances and later also over kilometre-scale distances in free space²⁰. However, in these experiments only two out of four Bell states were distinguished unambiguously, limiting the efficiency of the protocol to 50% at best. A proof-of-principle experiment that can distinguish all four Bell states was implemented using nonlinear photon interaction²¹, but the efficiency of the detection step was much below 1%. With photonic continuous-variable states, teleportation has been achieved deterministically for all measurement outcomes and the final conditional rotation has been implemented to complete the teleportation protocol²². In atomic qubits, fully deterministic quantum teleportation has been realized over micrometre-scale distances with ions in the same trap^{23,24}. Non-deterministically the protocol has also been implemented between ions in different traps²⁵ and in atomic ensembles²⁶.

The experimental realization of teleportation faces several important challenges that have been sequentially addressed over many years of research in most physical realizations. The ideal protocol requires the creation of an entangled pair shared between the sender and the receiver, a two-qubit measurement identifying all four Bell states at the sender, and feed-forward of the classical information to perform the final qubit state rotation to complete the protocol, all at high efficiency and fidelity. In addition it is frequently desired to run the protocol at a high rate and over large distances to maximize its usefulness.

Here we demonstrate in a solid state system post-selected teleportation, deterministic teleportation and deterministic teleportation with feed-forward in a single experimental setting. Our experiment addresses all of the above criteria with the exception of realizing space-like separation between the sender and the receiver. The coherent part of the teleportation protocol presented here has previously been characterized

in our laboratory by performing full-state tomography of three qubits interacting with a single mode field²⁷. That work has allowed us to test the protocol up to, but not including, the essential single-shot readout and feed-forward steps, which are the focus of this study.

In the realization of teleportation presented here, we use three superconducting transmon qubits³ (Q1, Q2 and Q3) coupled to three superconducting coplanar waveguide resonators (R1, R2 and R3) in a circuit quantum electrodynamics setup¹² (Fig. 2). At the sender, qubits Q1 and Q2 are coupled capacitively to resonator R1; at the receiver, Q3 is coupled to R3. The individual resonators R1 and R3 allow the sender and the receiver to perform independent measurements of their qubits, as detailed below. In addition, Q2 and Q3 are coupled to R2. The resonators R2 and R1 act as quantum buses to realize two-qubit controlled-PHASE (CPHASE) gates⁷ used to create a shared Bell state distributed between the sender and the receiver and to perform a deterministic Bell state analysis at the sender. We perform single-qubit rotations with an average fidelity of greater than 98% by applying amplitude-controlled and phase-controlled microwave pulses through individual charge gate lines. The transition frequency of each qubit is controlled by individual flux bias lines.

We start the protocol by heralding the three-qubit ground state based on a single-shot measurement of all qubits. Using single-qubit rotations and a CPHASE gate with process fidelity 90%, a Bell state with state fidelity 92% is created on demand between qubits Q2 and Q3 that is shared between the sender and the receiver (blue box in Fig. 1b). Only after the creation of the shared entanglement, we use a single-qubit rotation to prepare the state of qubit Q1 to be teleported.

As an essential part of the protocol, we perform a Bell measurement of qubits Q1 and Q2 at the sender by mapping the Bell basis $\{|\Phi^-\rangle, |\Psi^-\rangle, |\Phi^+\rangle, |\Psi^+\rangle\}$ onto the computational basis $\{|00\rangle, |01\rangle, |10\rangle, |11\rangle\}$ using a CNOT gate and a Hadamard gate (red box in

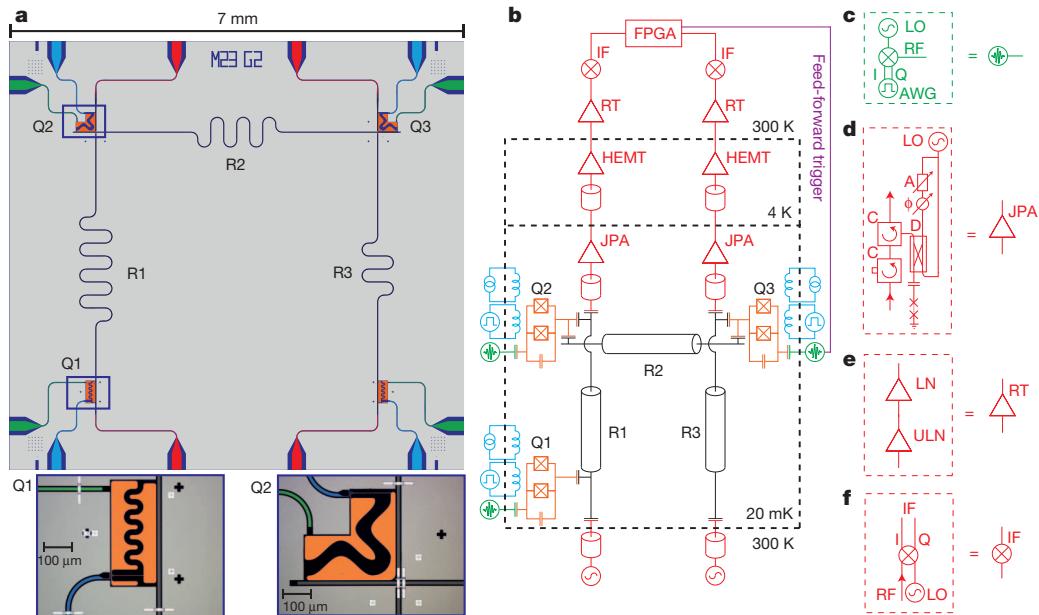


Figure 2 | Diagrams of sample and measurement setup. **a**, Chip design including three resonators R1, R2 and R3 (black) with corresponding input and output lines (red) used for readout and coupling of three transmon qubits Q1, Q2 and Q3 (orange). The fourth qubit in the lower right corner of the chip is not used. The local microwave charge gate lines (green) are used for single-qubit rotations; the local flux-bias lines (blue) permit nanosecond time control of the qubit frequencies to implement two-qubit operations. Below: false-colour micrographs of Q1 (left) coupled to resonator R1, and Q2 (right) coupled to both resonators R1 and R2. The microfabricated aluminium air bridges visible as bright white strips realize crossovers for the resonator lines. Air bridges are also used to suppress spurious electromagnetic modes by connecting the ground planes across the coplanar wave guides. **b**, Simplified diagram of the measurement setup with the same colour code as in **a**. **c**, Amplitude-controlled and phase-controlled microwave pulses are applied to the qubits using

sideband modulation of an up-conversion in-phase quadrature (IQ) mixer driven by a local oscillator (LO) and modulated by an arbitrary waveform generator (AWG). **d**, The measurement signals transmitted through R1 and R3 are amplified using Josephson parametric amplifiers (JPA) pumped by an LO through a directional coupler (D), which is also used for phase (ϕ) and amplitude (A) controlled cancellation of the pump leakage, and coupled to and isolated from the sample by two circulators (C). **e**, The signal is further amplified by high-electron-mobility transistor (HEMT) amplifiers at 4 K and chains of ultra-low-noise (ULN) and low-noise (LN) room-temperature (RT) amplifiers. **f**, The transmission signal of both resonators is down-converted to an intermediate frequency (IF) in an IQ mixer pumped by a dedicated LO, digitized, and fed into field-programmable gate array (FPGA) logic for real-time data analysis and triggering the conditioned feed-forward step.

Fig. 1a). We realize this basis transformation by using single-qubit rotations and a CPHASE gate. Then we perform a projective joint readout²⁸ of the states of Q1 and Q2 by measuring the transmission amplitude and phase of resonator R1. A given Bell state is transformed to the corresponding computational basis state, resulting in the corresponding output state $|\psi_{\text{out}}\rangle = \{\mathbb{I}, X, Z, \hat{Y}\} |\psi_{\text{in}}\rangle$ of Q3.

To perform a post-selected teleportation protocol, it is sufficient to distinguish only one of the four Bell states, say $|00\rangle$, ideally occurring with a probability of 1/4, with high fidelity from all other states ($|01\rangle, |10\rangle$ and $|11\rangle$) by using the aforementioned scheme. This is possible in our setup by performing a dispersive single-shot qubit readout with a fidelity of $(91.0 \pm 0.2)\%$. For this purpose we amplify a measurement tone applied to R1 with a Josephson parametric amplifier operated in the phase-sensitive mode, in which no or only very little noise is added to the signal⁹. If the measurement of Q1 and Q2 returns 00, qubit Q3 is instantaneously projected to the desired state $|\psi_{\text{in}}\rangle$ not requiring any additional rotations at the receiver's qubit to complete teleportation.

With a second, independent, parametric amplifier, also operated in the phase-sensitive mode, a measurement tone transmitted through resonator R3 is used to read out the state of qubit Q3 with a single-shot fidelity of $(89.1 \pm 0.5)\%$. In post-selection (ps), state tomography of Q3 conditioned on a 00 measurement of Q1 and Q2 reveals the original input state with an average fidelity of $\bar{\mathcal{F}}_{\text{s}}^{\text{ps},00} = (81.9 \pm 1.3)\%$ (Extended Data Fig. 3a). By characterizing $|\psi_{\text{out}}\rangle$ for four linearly independent input states $|\psi_{\text{in}}\rangle$, we perform full process tomography²⁹ of the state transfer by teleportation from Q1 to Q3 to reconstruct the process matrix χ_{00} , which has a fidelity of $\mathcal{F}_{\text{p}}^{\text{ps},00} = (72.5 \pm 1.3)\%$ with respect to the expected identity operation. The process fidelity \mathcal{F}_{p} is related to the average output-state fidelity $\bar{\mathcal{F}}_{\text{s}}$ by $\mathcal{F}_{\text{p}} = (\bar{\mathcal{F}}_{\text{s}}(d+1) - 1)/d$, where d is the dimensionality of the input state and output state.

It is important to note that in post-selection we are able to map any of the four Bell states to the computational basis state $|00\rangle$ on demand by applying π -pulses to Q1 and/or Q2 right before their joint readout. This allows us to post-select individually on any of the four Bell states and to determine the corresponding process matrices $\chi_{00,01,10,11}$. The experimentally obtained process matrices (Fig. 3a) agree well with the expected processes. The average output-state fidelity $\bar{\mathcal{F}}_{\text{s}}^{\text{ps}} = (81.7 \pm 1.4)\%$ of all four processes is clearly above the classical limit of 2/3 (ref. 18). This results in an average process fidelity when post-selecting on a single Bell state of $\bar{\mathcal{F}}_{\text{p}}^{\text{ps}} = (72.0 \pm 1.4)\%$, well above the classical limit of 1/2. The output-state fidelity is predominantly limited by the relaxation and dephasing of our qubits, which affects both the effective gate and readout fidelity (see Methods).

To identify uniquely and simultaneously the four randomly distributed outcomes of the Bell-state measurement at the sender in a deterministic way (instead of using post-selection) we use high-fidelity dispersive single-shot qubit readout enabled by the Josephson parametric amplifier operated in the phase-preserving mode³⁰. In this mode we achieve a probability of $(81.8 \pm 0.5)\%$ to correctly identify all four pairs of basis states, by amplifying and recording both quadrature amplitudes (I, Q) of the detected electromagnetic field at the cost of adding at least the vacuum noise to the input signal. Correlating the four single-shot Bell-state measurement outcomes at the sender with the single-shot measurement outcomes at the receiver, and performing state and process tomography, we find an average deterministic output-state fidelity of the transferred state of $\bar{\mathcal{F}}_{\text{s}}^{\text{det}} = (77.1 \pm 1.2)\%$ and an average process fidelity of $\bar{\mathcal{F}}_{\text{p}}^{\text{det}} = (65.5 \pm 1.1)\%$, well above the classical limits of 2/3 and 1/2, respectively. The process matrices (Fig. 3b) prominently show the characteristic features of the expected processes. The fidelities obtained with this method are lower than for the post-selected teleportation, because of the lower fidelity of the deterministic Bell-state readout.

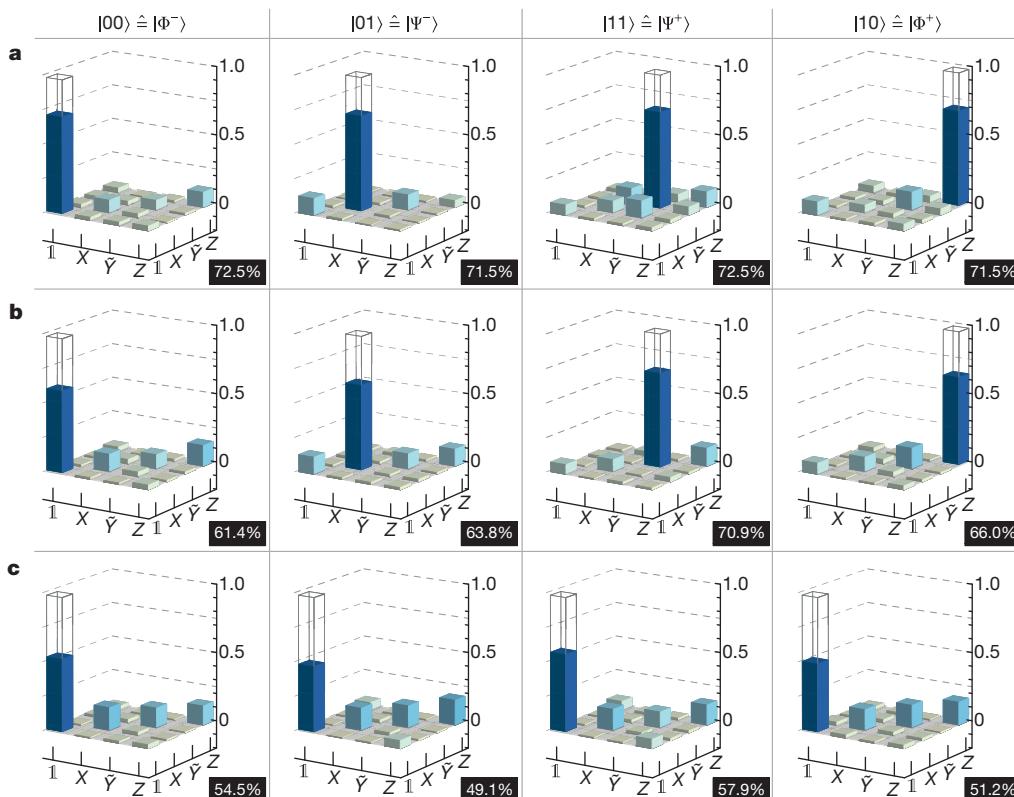


Figure 3 | State-transfer process matrix for quantum teleportation. The experimentally obtained (solid bars) absolute values of the process matrices $|\chi|$ describing the state transfer from Q1 to Q3 by teleportation are displayed for post-selection on any single one of the Bell-measurement outcomes 00, 01, 10

and 11 (a), simultaneous deterministic measurement of all four outcomes (b), and with feed-forward (c). The respective process fidelities are indicated in black boxes. The ideal $|\chi|$ is indicated by wire frames.

To complete the teleportation protocol we have analysed the deterministic Bell-state measurement in real time by using fast electronics based on a field programmable gate array (FPGA). Digitally defining two thresholds for the single-shot measurements of the quadrature amplitudes I and Q performed at the sender we realize the feed-forward (ff) step in about 500 ns, in which the rotations I , X , $\tilde{Y}X$ ($=Z$), \tilde{Y} are applied conditioned on the four Bell-state measurement outcomes 00, 01, 10 and 11. We analysed more than 270,000 single-shot experiments and find a deterministic state-transfer fidelity of $\mathcal{F}_s^{\text{ff}} = (68.8 \pm 0.5)\%$ with a process fidelity $\mathcal{F}_p^{\text{ff}} = (53.2 \pm 0.5)\%$, close to but clearly above the classical thresholds (Fig. 3c). The limited fidelity can be traced back to the fidelities of single-qubit and two-qubit operations, the readout fidelities and the time required for the feed-forward in relation to the coherence times of the qubits used.

When averaging and appropriately normalizing the measurement results of Q3 directly in the FPGA, we are able to circumvent the limited single-shot readout fidelity of Q3. In this case, the deterministically teleported states including the feed-forward step have an average fidelity of $\mathcal{F}_s^{\text{av}} = (77.4 \pm 0.2)\%$, with a corresponding process fidelity of $\mathcal{F}_p^{\text{av}} = (62.2 \pm 0.3)\%$.

To characterize the feed-forward process only, we have prepared separable input states ($|00\rangle$, $|01\rangle$, $|10\rangle$ and $|11\rangle$) of Q1 and Q2, and arbitrary states $|\psi\rangle$ of Q3. The input states measured and digitally analysed at the sender are then used to actuate the feed-forward. The fidelity of the resulting state of Q3 subject to feed-forward was determined to be $\mathcal{F}_s = (79.5 \pm 1.5)\%$ with a corresponding process fidelity of $\mathcal{F}_p = (72.4 \pm 1.5)\%$. These results also indicate that this type of deterministic feed-forward can in principle be used for demonstrating error correction in future experiments, if high enough fidelities can be realized.

These results, in conjunction with continuing improvements in coherence time³, advances in circuit design and operation, and the potential for using propagating microwave photons, form a solid ground for future progress in quantum information processing and quantum communication with superconducting circuits.

METHODS SUMMARY

The teleportation protocol (Fig. 1b) is implemented using the pulse scheme depicted in Extended Data Fig. 1. Single-qubit rotations are realized by 12 ns long resonant microwave pulses. Controlled-PHASE gates are implemented by shifting one of the two qubits' transition frequencies to the avoided level crossing between the $|11\rangle$ and $|02\rangle$ states. For single-shot measurements, the output signals of resonators R1 and R3 are amplified using Josephson parametric amplifiers. For the experiments in which we post-select on an individual Bell state, we operate the parametric amplifier in the phase-sensitive mode. This allows us to distinguish the state $|00\rangle$ from all other states by analysing the integrated measurement signal (Extended Data Fig. 2a). In the experiments in which we simultaneously distinguish between all four Bell states, the parametric amplifier is used in the phase-preserving mode, in which both quadratures of the transmitted field are amplified. By adjusting the pump and readout powers we maximize the distinguishability of all four states by their respective quadrature amplitudes in the complex plane (Extended Data Fig. 2b). To implement the feed-forward, all measurement data are analysed in real-time in an FPGA with custom firmware. Depending on the measurement outcome, the FPGA triggers up to two arbitrary waveform generators, which realize X and Y rotations. We characterize the state transfer from Q1 to Q3 with full process tomography by performing state tomography on the output state for four different linearly independent input states (Extended Data Fig. 3). We also use the teleportation protocol to verify the entanglement between Q2 and Q3 using six mutually unbiased input basis states.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 May; accepted 28 June 2013.

- Clarke, J. & Wilhelm, F. K. Superconducting quantum bits. *Nature* **453**, 1031–1042 (2008).

- Houck, A. A., Tureci, H. E. & Koch, J. On-chip quantum simulation with superconducting circuits. *Nature Phys.* **8**, 292–299 (2012).
- Devoret, M. & Schoelkopf, R. J. Superconducting circuits for quantum information: an outlook. *Science* **339**, 1169–1174 (2013).
- Steffen, M. et al. Measurement of the entanglement of two superconducting qubits via state tomography. *Science* **313**, 1423–1425 (2006).
- DiCarlo, L. et al. Preparation and measurement of three-qubit entanglement in a superconducting circuit. *Nature* **467**, 574–578 (2010).
- Neeley, M. et al. Generation of three-qubit entangled states using superconducting phase qubits. *Nature* **467**, 570–573 (2010).
- DiCarlo, L. et al. Demonstration of two-qubit algorithms with a superconducting quantum processor. *Nature* **460**, 240–244 (2009).
- Reed, M. D. et al. Realization of three-qubit quantum error correction with superconducting circuits. *Nature* **482**, 382–385 (2012).
- Castellanos-Beltran, M. A., Irwin, K. D., Hilton, G. C., Vale, L. R. & Lehnert, K. W. Amplification and squeezing of quantum noise with a tunable Josephson metamaterial. *Nature Phys.* **4**, 929–931 (2008).
- Vijay, R. et al. Stabilizing Rabi oscillations in a superconducting qubit using quantum feedback. *Nature* **490**, 77–80 (2012).
- Johnson, J. E. et al. Heralded state preparation in a superconducting qubit. *Phys. Rev. Lett.* **109**, 050506 (2012).
- Wallraff, A. et al. Strong coupling of a single photon to a superconducting qubit using circuit quantum electrodynamics. *Nature* **431**, 162–167 (2004).
- Helmer, F. et al. Cavity grid for scalable quantum computation with superconducting circuits. *Europhys. Lett.* **85**, 50007 (2009).
- Mariantoni, M. et al. Implementing the quantum von Neumann architecture with superconducting circuits. *Science* **334**, 61–65 (2011).
- Bennett, C. H. et al. Teleporting an unknown quantum state via dual classical and Einstein–Podolsky–Rosen channels. *Phys. Rev. Lett.* **70**, 1895–1899 (1993).
- Gisin, N., Ribordy, G., Tittel, W. & Zbinden, H. Quantum cryptography. *Rev. Mod. Phys.* **74**, 145–195 (2002).
- Gottesman, D. & Chuang, I. L. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature* **402**, 390–393 (1999).
- Massar, S. & Popescu, S. Optimal extraction of information from finite quantum ensembles. *Phys. Rev. Lett.* **74**, 1259–1263 (1995).
- Bouwmeester, D. et al. Experimental quantum teleportation. *Nature* **390**, 575–579 (1997).
- Ma, X.-S. et al. Quantum teleportation over 143 kilometres using active feed-forward. *Nature* **489**, 269–273 (2012).
- Kim, Y. H., Kulik, S. P. & Shih, Y. Quantum teleportation of a polarization state with a complete Bell state measurement. *Phys. Rev. Lett.* **86**, 1370–1373 (2001).
- Furusawa, A. et al. Unconditional quantum teleportation. *Science* **282**, 706–709 (1998).
- Riebe, M. et al. Deterministic quantum teleportation with atoms. *Nature* **429**, 734–737 (2004).
- Barrett, M. et al. Deterministic quantum teleportation of atomic qubits. *Nature* **429**, 737–739 (2004).
- Olmschenk, S. et al. Quantum teleportation between distant matter qubits. *Science* **323**, 486–489 (2009).
- Bao, X.-H. et al. Quantum teleportation between remote atomic-ensemble quantum memories. *Proc. Natl. Acad. Sci. USA* **109**, 20347–20351 (2012).
- Baur, M. et al. Benchmarking a quantum teleportation protocol in superconducting circuits using tomography and an entanglement witness. *Phys. Rev. Lett.* **108**, 040502 (2012).
- Filipp, S. et al. Two-qubit state tomography using a joint dispersive readout. *Phys. Rev. Lett.* **102**, 200402 (2009).
- Chuang, I. L. & Nielsen, M. A. Prescription for experimental determination of the dynamics of a quantum black box. *J. Mod. Opt.* **44**, 2455–2467 (1997).
- Eichler, C. et al. Observation of two-mode squeezing in the microwave frequency domain. *Phys. Rev. Lett.* **107**, 113601 (2011).

Acknowledgements We thank A. Blais, F. Marquardt, S. Filipp and R. Renner for discussion and feedback; and J. Heinsoo, L. Heinzel, A. Landig, Y. Liu, F. Lüthi and T. Menke for technical contributions to the experimental work. This work was supported financially by Eidgenössische Technische Hochschule Zurich (ETH Zurich), the EU Integrated Projects SOLID and SCALEQIT, and the Swiss National Science Foundation's National Centre of Competence in Research 'Quantum Science & Technology'.

Author Contributions The experiments were performed by L.S., A.F., Y.S., M.O., P.K. and M.B. The teleportation sample and parametric amplifiers were fabricated by A.F., L.S. and M.B. The air-bridge technology was developed by G.P.-H. and L.S. The FPGA firmware was implemented by Y.S. and C.L. The parametric amplifiers were designed by C.E., who also oversaw their operation. The manuscript was written by L.S., A.F. and A.W. All authors commented on the manuscript. The project was led by A.F. and A.W.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L.S. (isteffen@phys.ethz.ch) or A.W. (andreas.wallraff@phys.ethz.ch).

METHODS

Sample parameters. The sample consists of three superconducting coplanar waveguide resonators and three qubits of the transmon type³¹ as depicted in Fig. 2. The resonators R1 and R3 have bare resonance frequencies $v_r = \{7.657, 9.677\}$ GHz, respectively. They are coupled by gap and finger capacitors to their input and output lines. The overcoupled resonator decay rates are measured to be $\kappa/2\pi = \{2.4, 2.5\}$ MHz. The coupling capacitances are designed asymmetrically such that the decay rate through the input port is approximately 100-fold lower than through the output port. The resonator R2 is not coupled to any input or output line. Its resonance frequency is about 8.7 GHz and its decay rate is expected to be close to the internal decay rate³². From spectroscopic measurements we determine the maximum transition frequencies $v_{\max} = \{6.273, 7.373, 8.390\}$ GHz and charging energies $E_C/h = \{0.297, 0.303, 0.287\}$ GHz of the qubits Q1, Q2 and Q3, respectively, where h is Planck's constant.

Qubits Q1 and Q2 are coupled to resonator R1 with coupling strengths $g/2\pi = \{0.260, 0.180\}$ GHz, and Q3 is coupled to resonator R3 with a coupling strength of $g/2\pi = 0.240$ GHz. The coupling of Q2 and Q3 to R2 is estimated from the transverse coupling strength (see below) to be $g/2\pi = 0.2$ GHz each.

For the presented experiments, the qubits were tuned to idle state transition frequencies $v = \{4.657, 5.390, 6.342\}$ GHz with miniature superconducting coils mounted underneath the chip³³. At these frequencies we have determined their energy relaxation $T_1 = \{5.1, 3.1, 2.6\}$ μ s and coherence times $T_2 = \{1.0, 1.1, 1.7\}$ μ s. **Pulse scheme.** All biased qubits and resonators are separated in frequency from each other by at least 750 MHz to suppress cross-talk.

The protocol (Fig. 1b) is implemented with the pulse scheme depicted in Extended Data Fig. 1. Single-qubit rotations are implemented by resonant Gaussian-shaped DRAG^{34,35} microwave pulses 12 ns long with an average fidelity of more than 98% as determined by randomized benchmarking³⁶. The controlled-PHASE gate is implemented by shifting the qubits with fast magnetic flux pulses to the avoided level crossing between the $|11\rangle$ and $|02\rangle$ states of the involved qubits^{7,37}. The transverse coupling strengths of $J_{11,01}^{Q1,Q2}/2\pi = 12.2$ MHz (between qubits Q1 and Q2) and $J_{11,02}^{Q2,Q3}/2\pi = 10.4$ MHz (between Q2 and Q3) lead to pulse lengths for the CPHASE gates of $t = \{40.9, 47.9\}$ ns, respectively.

The Bell measurement (Extended Data Fig. 1, red elements) allows us to map any of the four Bell states to the $|00\rangle$ state by adding π -pulses to Q1 and Q2 to flip their states just before starting the measurement. Instead of applying these π -pulses directly, we change the phases of the preceding $\pi/2$ -pulses accordingly, which can easily be verified to be equivalent.

Qubit readout. To realize single-shot measurements, the output signals of resonators R1 and R3 are amplified by dedicated Josephson parametric amplifiers^{9,38}. The parametric amplifiers are similar to that used in ref. 30. They are realized as $\lambda/4$ coplanar waveguide resonators terminated by an array of 11 superconducting quantum interference devices (SQUIDS), which provide the necessary nonlinearity and make the operation frequencies tunable by miniature superconducting coils on the bottom of the sample holder³⁹. The maximum frequencies for the two parametric amplifiers are $v_{\max} = \{8.349, 10.141\}$ GHz for R1 and R3, respectively. To provide a fast response, large input capacitors were fabricated that resulted in a measured Josephson parametric amplifier line width of $\kappa/2\pi = \{334, 548\}$ MHz in the linear regime. For the experiments the parametric amplifiers were tuned to have a maximum gain of $G = \{20.4, 20.6\}$ dB with a 3-dB bandwidth of $B/2\pi = \{16, 85\}$ MHz at frequencies $v_{\exp} = \{7.686, 9.737\}$ GHz.

For the experiments in which we post-selected on an individual Bell state, the transmission of R1 was measured at the readout frequency $v_{ro} = 7.686$ GHz, which is the mean value of the effective resonator frequencies for the qubits Q1 and Q2 in the state $|00\rangle$ and $|01\rangle$. The parametric amplifier is used in the phase-sensitive mode by tuning its transition frequency such that the maximum gain was achieved at the readout frequency v_{ro} at which it was also pumped. Preparing the four computational basis states $|00\rangle, |01\rangle, |10\rangle$ and $|11\rangle$, applying a measurement tone to R1 and integrating the amplified transmission signal for 280 ns resulted in a distribution of the integrated measurement signals as shown in Extended Data Fig. 2a. We optimized for the readout contrast between the $|00\rangle$ and all other states. In this way, the mean values of the distributions of the integrated signals for the states $|00\rangle, |10\rangle$ (Extended Data Fig. 2a) have the largest separation. However, as a result of the finite qubit lifetime, some of the $|01\rangle$ and $|10\rangle$ states decay into the ground state and are visible in the data as such. We choose a threshold for the integrated quadrature values to discriminate 00 from all other measurement outcomes 01, 10 and 11 with a fidelity of $91.0 \pm 0.2\%$.

In the experiments in which we are able to distinguish between all four Bell states simultaneously, the readout frequency $v_{ro} = 7.683$ GHz is chosen to be the mean of the effective resonator frequencies for the qubits Q1 and Q2 in the state $|01\rangle$ and $|10\rangle$. The parametric amplifier is used in the phase-preserving mode by detuning the pump frequency by 6.25 MHz from the readout frequency. In this way the gain $G = 18.4$ dB at the readout frequency and the effective bandwidth are

smaller than for phase-sensitive amplification, but both quadratures of the transmitted field are amplified. By preparing the computational basis states and recording the integrated transmitted signals of both quadratures (I, Q) simultaneously, we can map every measurement outcome to a point on the complex plane. By adjusting the pump power and the readout power we find settings that maximize the distinguishability of all four states by their location in the complex plane.

Adjusting the phase of the local oscillator and implementing small linear offsets directly in the FPGA we are able to choose the $I = 0$ and $Q = 0$ axes of the complex plane as thresholds to identify the four different output states in real time (Extended Data Fig. 2b). By assigning our best estimate of the corresponding states to every measurement outcome according to a quadrant in the I - Q plane, we identify $(81.8 \pm 0.5)\%$ of the prepared states correctly. The probabilities of either correctly identifying a prepared state or misidentifying it as a different state are listed in Extended Data Table 1 for all four input basis states.

Feed-forward. To implement the feed-forward, all measurement data are analysed in real-time in an FPGA with custom firmware. Depending on the measurement outcome, the FPGA triggers up to two different arbitrary waveform generators (AWGs), which realize X and \tilde{Y} rotations. To optimize the readout fidelity and minimize the time needed for data analysis, we calibrate the phase of the down-conversion local oscillator to rotate the readout data of Q1 and Q2 such that the four different states can be distinguished by thresholds parallel to the I and Q axes of the complex plane. An additional small offset is subtracted digitally such that the assignment of a measured signal to a given state is based solely on the quadrant of the complex plane it lies in (see Extended Data Fig. 2b).

We integrate the signal for 160 ns to achieve the desired signal-to-noise ratio. This choice of integration time also digitally filters out the parametric amplifier pump tone, which is detuned by 6.25 MHz from the measurement tone. We experimentally determine a maximum readout fidelity by starting the integration time 140 ns after the initial rise of the pulsed measurement tone applied to the resonator. The FPGA used for data analysis and the AWG used for generating the feed-forward pulses introduce a combined delay of 140 ns; the total cable delay is 65 ns. As a result, the conditional rotations are applied to Q3 505 ns after the measurement tone has been applied to the readout resonator R1 (Extended Data Fig. 1). To mitigate dephasing of Q3 during this feed-forward delay time, we apply a series of four dynamical decoupling pulses to Q3.

We have also fully characterized the fidelity of the feed-forward process. After preparing the state indicated in the rows of Extended Data Table 2 and performing the conditional rotations indicated in the columns of Extended Data Table 2, the fidelity of the applied feed-forward process was determined. The diagonal elements indicate the fidelity of the desired feed-forward operations occurring with the probabilities shown in Extended Data Table 1. The off-diagonal elements indicate the fidelity of the unwanted rotations occurring with a smaller probability (Extended Data Table 1) but a similar fidelity.

Efficiency of the teleportation protocol. The total efficiency of the implemented teleportation protocol is close to unity. The main limitation is currently set by the heralding process of the initial ground state of all qubits. Before each individual experimental realization of the teleportation protocol, we apply a measurement tone 500 ns long to both resonators, to verify that all the qubits are in their ground states. This heralds the ground state with a total efficiency of more than 80% when using the parametric amplifiers in the phase-sensitive mode. In the phase-preserving mode we herald the ground state with $\sim 65\%$ probability while performing single-shot readout. This is in stark contrast to experiments in which optical photons are used as qubits, where the maximal reported efficiency is 0.1% (ref. 40). Using active initialization schemes for superconducting circuits^{41–45}, the efficiency of future experiments can probably be improved to approach 100% for sufficiently long qubit coherence times.

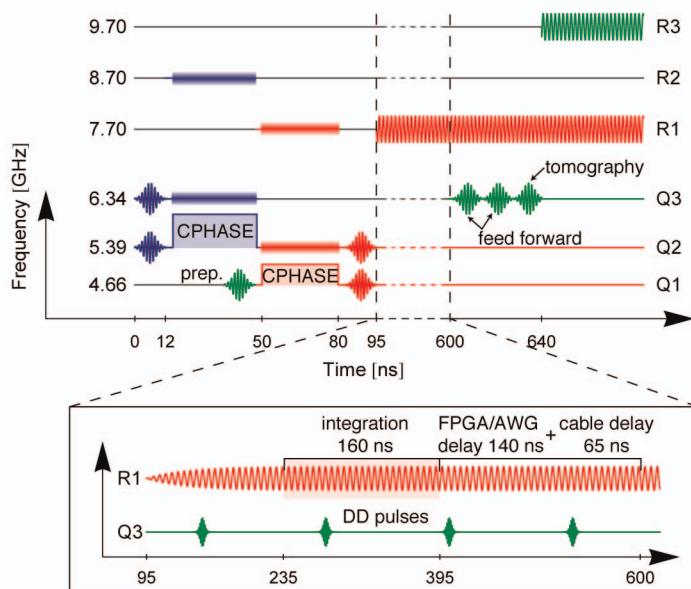
State and process tomography. To characterize the state transfer from Q1 to Q3 we performed full process tomography²⁹. Performing state tomography on the output state $|\psi_{\text{out}}\rangle$ for four different input states $|\psi_{\text{in}}\rangle = |0\rangle, |1\rangle, (|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle - i|1\rangle)/\sqrt{2}$ (Extended Data Fig. 3), we obtain the process matrix χ through linear inversion. In state tomography we measure a state prepared repeatedly with a set of orthogonal measurement operators realized by applying either no pulse, a $\pi/2$ -pulse about the x or y axis or a π -pulse to the qubit before measurement of its Z eigenstate.

Entanglement verification in a teleportation protocol. The initial shared two-qubit state was characterized by performing state tomography immediately after its generation, instead of continuing the teleportation protocol. The resulting Bell state has a fidelity of 92% and a concurrence of 0.89, demonstrating the entangled nature of the state. In addition it is also possible to use the teleportation protocol itself as method for entanglement verification⁴⁶. If the average output-state fidelity for six mutually unbiased input basis states is above the classical threshold of $2/3$, it unambiguously shows that the two parties shared an entangled state, even if one has no knowledge about the details of the protocol itself. We performed the

teleportation experiment with feed-forward also with the six mutually unbiased input basis states $|\psi_{in}\rangle = |0\rangle, |1\rangle, (|0\rangle + |1\rangle)/\sqrt{2}, (|0\rangle - i|1\rangle)/\sqrt{2}, (|0\rangle - |1\rangle)/\sqrt{2}$ and $(|0\rangle + i|1\rangle)/\sqrt{2}$ and found an average output-state fidelity $\mathcal{F}_s^{ff} = (69.3 \pm 0.1)\%$ with a process fidelity $\mathcal{F}_p^{ff} = (53.9 \pm 0.1)\%$.

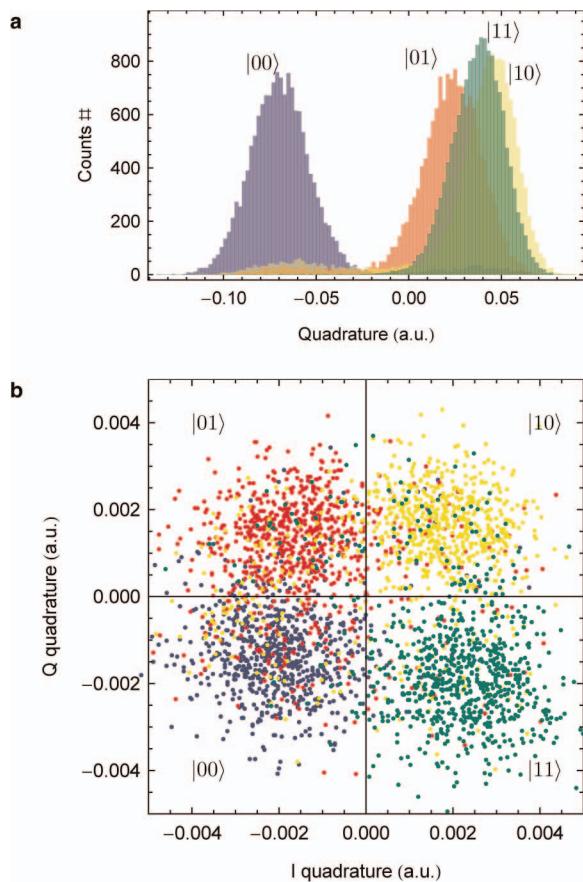
Error budget. The finite coherence and dephasing times of our qubits are a source of error that limits the output-state fidelity. The fidelity of the measurement of Q3 through R3 affects the state fidelity of $|\psi_{out}\rangle$ directly. From the measured probabilities of correctly identifying the states $|0\rangle$ and $|1\rangle$ on Q3 we calculate the limit of the output-state fidelity through this source of error to be $\mathcal{F}_s = 94\%$. In addition, the misidentification of the Bell states of Q1 and Q2 leads to an effective dephasing of $|\psi_{out}\rangle$. This limits the fidelity further to $\mathcal{F}_s = 89\%$ and $\mathcal{F}_s = 84\%$ for the respective cases in which we post-select on one Bell state only and in which we distinguish all Bell states with each measurement. Because both of these numbers are about 7–8% higher than the actually measured fidelities, it is plausible to assign the remaining errors to the limited gate fidelities. Determining the gate errors independently shows that we perform single-qubit operations with a fidelity greater than 98% and create Bell states with a fidelity of 92% using two-qubit CPHASE gates with a process fidelity of 90%.

31. Koch, J. *et al.* Charge-insensitive qubit design derived from the Cooper pair box. *Phys. Rev. A* **76**, 042319 (2007).
32. Göppl, M. *et al.* Coplanar waveguide resonators for circuit quantum electrodynamics. *J. Appl. Phys.* **104**, 113904 (2008).
33. Fink, J. M. *et al.* Dressed collective qubit states and the Tavis–Cummings model in circuit QED. *Phys. Rev. Lett.* **103**, 083601–083604 (2009).
34. Motzoi, F., Gambetta, J. M., Rebentrost, P. & Wilhelm, F. K. Simple pulses for elimination of leakage in weakly nonlinear qubits. *Phys. Rev. Lett.* **103**, 110501 (2009).
35. Gambetta, J. M., Motzoi, F., Merkel, S. T. & Wilhelm, F. K. Analytic control methods for high-fidelity unitary operations in a weakly nonlinear oscillator. *Phys. Rev. A* **83**, 012308 (2011).
36. Chow, J. M. *et al.* Randomized benchmarking and process tomography for gate errors in a solid-state qubit. *Phys. Rev. Lett.* **102**, 090502 (2009).
37. Strauch, F. W. *et al.* Quantum logic gates for coupled superconducting phase qubits. *Phys. Rev. Lett.* **91**, 167005 (2003).
38. Vijay, R., Slichter, D. H. & Siddiqi, I. Observation of quantum jumps in a superconducting artificial atom. *Phys. Rev. Lett.* **106**, 110502 (2011).
39. Eichler, C. & Wallraff, A. Controlling the dynamic range of a Josephson parametric amplifier. *arXiv:1305.6583* (2013).
40. Nölleke, C. *et al.* Efficient teleportation between remote single-atom quantum memories. *Phys. Rev. Lett.* **110**, 140403 (2013).
41. Ristè, D., van Leeuwen, J. G., Ku, H.-S., Lehnert, K. W. & DiCarlo, L. Initialization by measurement of a superconducting quantum bit circuit. *Phys. Rev. Lett.* **109**, 050507 (2012).
42. Ristè, D., Bultink, C. C., Lehnert, K. W. & DiCarlo, L. Feedback control of a solid-state qubit using high-fidelity projective measurement. *Phys. Rev. Lett.* **109**, 240502 (2012).
43. Murch, K. W. *et al.* Cavity-assisted quantum bath engineering. *Phys. Rev. Lett.* **109**, 183602 (2012).
44. Geerlings, K. *et al.* Demonstrating a driven reset protocol for a superconducting qubit. *Phys. Rev. Lett.* **110**, 120501 (2013).
45. Campagne-Ibarcq, P. *et al.* Persistent control of a superconducting qubit by stroboscopic measurement feedback. *Phys. Rev. X* **3**, 021008 (2013).
46. van Enk, S. J., Lütkenhaus, N. & Kimble, H. J. Experimental procedures for entanglement verification. *Phys. Rev. A* **75**, 052318 (2007).

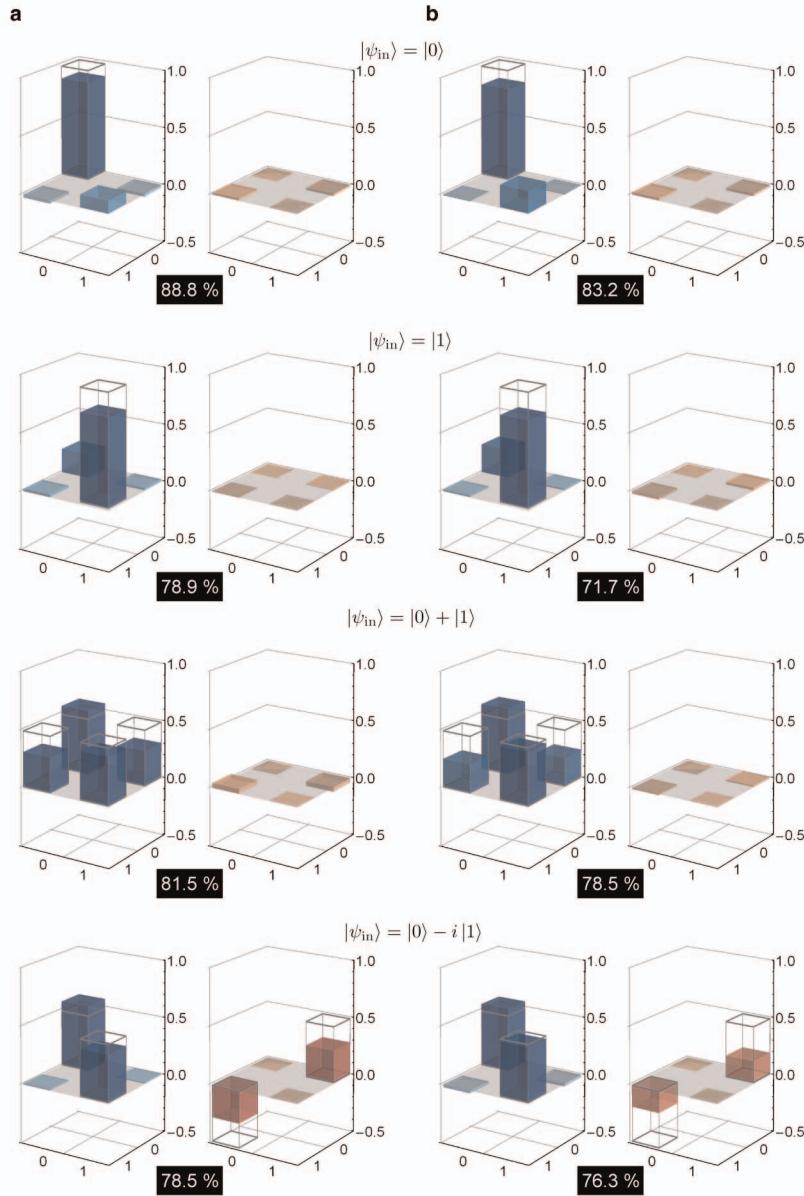


Extended Data Figure 1 | Pulse sequence of the teleportation protocol with feed-forward. The pulses implement the creation of an entangled pair between Q2 and Q3 (blue), the preparation of the state to be teleported on Q1 (green), the basis transformation from the Bell to the computational basis and the subsequent readout of Q1 and Q2 (red), the dynamical decoupling (DD) pulses, conditional rotations and the state tomography on Q3 (green). Gaussian-shaped sinusoids represent the microwave pulses applied to the respective charge bias lines of the qubits; sinusoids on the resonators represent the readout

tones; and the squares labelled CPHASE represent the flux pulses that shift the frequency of a qubit to implement a controlled-PHASE gate between the marked qubits, where the interaction is mediated through the resonator indicated with a bar of the same colour as the flux pulse. The inset shows the time used for implementing the conditional feed-forward rotations. The total feed-forward time is the sum of the ramp-up time of the measurement tone, the integration time of the measurement signal and the delay times induced by the FPGA signal processing, the AWG trigger and the cables.



Extended Data Figure 2 | Characterization of the joint readout of Q1 and Q2. **a**, Histogram of the integrated signal quadrature-amplitude amplified phase-sensitively when preparing the states $|00\rangle$ (blue), $|01\rangle$ (red), $|10\rangle$ (yellow) and $|11\rangle$ (green). **b**, Scatter plot of integrated (I , Q) quadratures of the measurement signal amplified in the phase-preserving mode when preparing the states $|00\rangle$ (blue), $|01\rangle$ (red), $|10\rangle$ (yellow) and $|11\rangle$ (green).



Extended Data Figure 3 | Characterization of the output states. Real (blue) and imaginary (red) parts of the reconstructed density matrices of the state $|\psi_{\text{out}}\rangle$ for the indicated input states $|\psi_{\text{in}}\rangle$ obtained from state tomography when (a) post-selecting data on a 00 outcome of the Bell measurement (b) using

averaged readout on Q3 while performing fully deterministic teleportation with feed-forward. The ideally expected outcomes are indicated with wireframes. The state fidelities are indicated in the black boxes.

Extended Data Table 1 | Success probabilities for the joint readout

	00	01	10	11
00>	0.88	0.09	0.01	0.02
01>	0.11	0.79	0.08	0.02
10>	0.06	0.10	0.77	0.06
11>	0.02	0.03	0.08	0.87

Probabilities of identifying prepared input states (rows) as the indicated output state (columns). Correct identifications are diagonal elements; misidentifications are off-diagonal elements.

Extended Data Table 2 | Process fidelities of the feed-forward pulses

	I	X	Y	Z
$ 00\rangle \otimes \psi\rangle$	0.80	0.67	0.74	0.64
$ 01\rangle \otimes X \psi\rangle$	0.64	0.69	0.73	0.65
$ 10\rangle \otimes Z \psi\rangle$	0.66	0.63	0.66	0.65
$ 11\rangle \otimes \tilde{Y} \psi\rangle$	0.73	0.68	0.62	0.75

After preparing the indicated input states (rows) and performing the conditional operations (columns), the fidelity of the expected process is determined (elements of the table). The rows indicate the fidelity of the performed rotation conditioned on the measurement result of Q1 and Q2, which occur with the probabilities shown in Extended Data Table 1.

Tunable near-infrared and visible-light transmittance in nanocrystal-in-glass composites

Anna Llordés¹, Guillermo Garcia¹, Jaume Gazquez² & Delia J. Milliron¹

Amorphous metal oxides are useful in optical^{1,2}, electronic^{3–5} and electrochemical devices^{6,7}. The bonding arrangement within these glasses largely determines their properties, yet it remains a challenge to manipulate their structures in a controlled manner. Recently, we developed synthetic protocols for incorporating nanocrystals that are covalently bonded into amorphous materials^{8,9}. This ‘nanocrystal-in-glass’ approach not only combines two functional components in one material, but also the covalent link enables us to manipulate the glass structure to change its properties. Here we illustrate the power of this approach by introducing tin-doped indium oxide nanocrystals into niobium oxide glass (NbO_x), and realize a new amorphous structure as a consequence of linking it to the nanocrystals. The resulting material demonstrates a previously unrealized optical switching behaviour that will enable the dynamic control of solar radiation transmittance through windows. These transparent films can block near-infrared and visible light selectively and independently by varying the applied electrochemical voltage over a range of 2.5 volts. We also show that the reconstructed NbO_x glass has superior properties—its optical contrast is enhanced fivefold and it has excellent electrochemical stability, with 96 per cent of charge capacity retained after 2,000 cycles.

Enhanced properties have been demonstrated when nanocrystalline precipitates form within a glassy matrix^{10–15}. These crystalline domains can harden bulk glasses¹⁰, introduce optical nonlinearity¹¹ or lead to anomalous transparency¹². Anecdotal evidence also suggests that ion transport may be enhanced at crystal–glass interfaces. For instance, when LiAlSiO_4 was partially crystallized, fast relaxation of ions was observed¹³, whereas precipitating TiO_2 nanocrystals within mesoporous P_2O_5 glass led to high-power lithium (Li) ion battery electrodes¹⁴. Structural modifications at crystal–glass interfaces have been proposed to explain such observations, but no direct evidence has been reported owing to the limited structural and compositional control offered by those preparation methods. Conventional methods of preparing glass ceramics rely on thermally induced phase separation and *in situ* crystallization processes, which are very complex to control experimentally¹⁶. The time–temperature parameter space for selective nanocrystallization is typically narrow and strongly influenced by the glass composition, structure and thermal history¹⁶. Therefore, determining the heat-treatment conditions required for selective crystallization of a desired composition phase with high control over the interface area becomes not only a challenging and tedious task, but also a nearly impossible task for complex multicomponent stoichiometries (for example, doped nanocrystals). Here we show how to control these characteristics by using well-defined colloidal nanocrystals that are post-synthetically embedded in and chemically linked to glass.

In our colloidal approach, chemical bonds are made by combining ligand-stripped Sn-doped In_2O_3 (ITO) nanocrystals^{17,18} and polyniobate clusters (POMs) in solution⁹. This process results in POM-stabilized nanocrystal dispersions (Fig. 1a, Supplementary Fig. 2), from which films are deposited and then thermally annealed in air, at $T = 400^\circ\text{C}$ (Fig. 1b). X-ray diffraction (XRD) analysis of as-deposited and annealed

films shows that the crystalline molecular POMs condense to an amorphous NbO_x matrix (Supplementary Fig. 3). In addition, XRD patterns (Fig. 1c, d) and scanning transmission electron microscopy cross-sectional images (Fig. 1e, f) confirm the amorphous nature of the NbO_x matrix and that the embedded ITO nanocrystals remain highly crystalline. This synthetic approach is very flexible, permitting the selection of unusual combinations of components to tailor functionality. For instance, ITO nanocrystals and amorphous NbO_x were specifically selected for their spectrally distinct electrochromic response. The volume fraction was finely tuned over a broad range, from 0% to 69%, by simply adding additional POM to the aqueous POM-stabilized nanocrystal dispersion (Fig. 1a). This tunability is key to revealing the reconstruction of the NbO_x glass structure induced by its covalent linkage to ITO nanocrystals.

To evaluate the structure of the glass, the Raman spectrum of a pure amorphous NbO_x film was compared to those of ITO-in- NbO_x with different nanocrystal–glass interfacial areas (Fig. 2). Vibrational bands of the distorted $[\text{NbO}_6]$ octahedra that dominate the structure of NbO_x

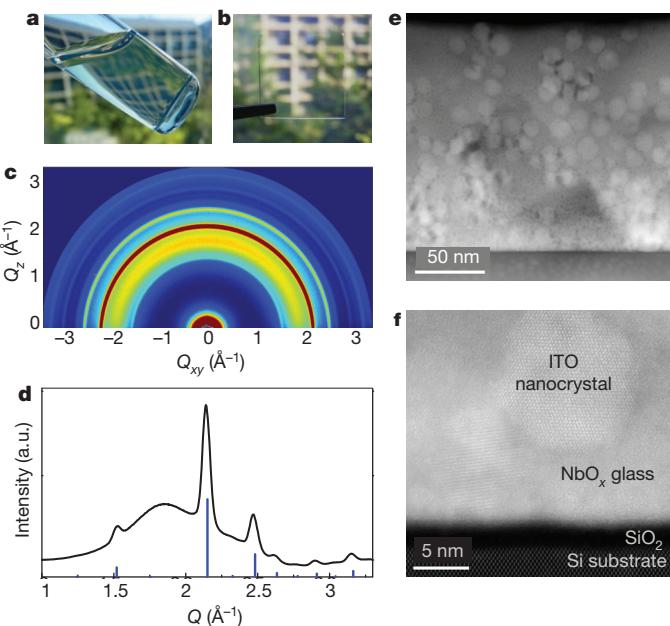


Figure 1 | Nanocrystal-in-glass film preparation and structural characterization. **a**, POM-stabilized ITO aqueous dispersion, that is, $[\text{N}(\text{CH}_3)_4]_6\text{Nb}_{10}\text{O}_{28}$. **b**, ITO-in- NbO_x film on a glass substrate. On annealing, the anionic $[\text{Nb}_{10}\text{O}_{28}]^{6-}$ clusters condense (Supplementary Fig. 3) and the $[\text{N}(\text{CH}_3)_4]^+$ counterions thermally decompose⁹. **c**, Two-dimensional XRD image of a ITO-in- NbO_x film. **d**, XRD pattern obtained after integrating along the out-of-plane scattering direction Q_z . The indexed diffraction lines result from randomly oriented ITO nanocrystals (cubic In_2O_3 bixbyite structure) whereas the broad peak at $Q = 1.85 \text{ \AA}^{-1}$ is assigned to amorphous NbO_x . **e, f**, Low- and high-resolution cross-sectional scanning transmission electron microscopy Z-contrast images (see Methods).

¹The Molecular Foundry, Lawrence Berkeley National Laboratory, 1 Cyclotron Road, Berkeley, California 94720, USA. ²Institut de Ciència de Materials de Barcelona, ICMAB-CSIC, Campus de la UAB, 08193 Bellaterra, Catalonia, Spain.

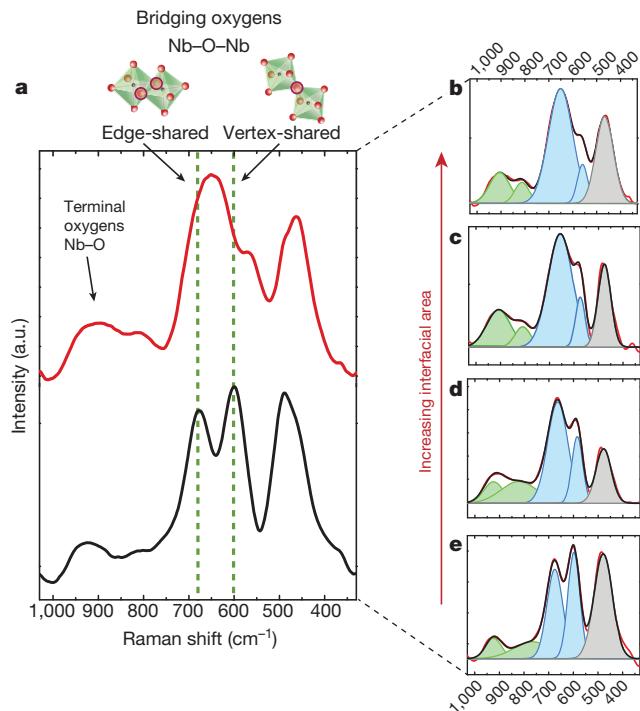


Figure 2 | Raman analysis probing the reconstruction of a NbO_x glass matrix when linked to nanocrystals. **a**, Raman spectra of pure NbO_x (black) and ITO-in- NbO_x nanocomposite (red) with 64 vol.% ITO. **b–e**, Fitted Raman spectra (experimental data in red, sum of fits in black). From bottom to top are pure NbO_x followed by nanocomposites with increasing ITO volume fraction (0.48, 0.52 and 0.64). Deconvolved peaks (grey) arise non-specifically from ITO, NbO_x and from the underlying glass substrate. Blue and green peaks specifically arise from NbO_x (bridging oxygens and terminal oxygens, respectively).

are readily observed in the range 500 cm^{-1} to $1,000 \text{ cm}^{-1}$ and changes in their peak positions and relative intensities track with the interfacial area. This information was used to identify the structural changes in the glass as a result of linking to the nanocrystals.

The topological structure of NbO_x glass is characterized by how the $[\text{NbO}_6]$ octahedral units are connected, namely through vertices and along edges, as well as by the degree of connectivity in the disordered network. Given the higher electrostatic repulsions in edge-sharing topologies compared to vertex-sharing¹⁹, distinctive vibrational modes for the bridging oxygens can be observed²⁰. By deconvolving these bridging oxygen bands (550 cm^{-1} to 750 cm^{-1}), we find that highly distorted edge-sharing $[\text{NbO}_6]$ are prevalent in the nanocrystal-in-glass films, whereas pure NbO_x contains more regular, vertex-sharing arrangements (Fig. 2, Table 1, Supplementary Fig. 4). In addition, the relative intensity of Raman peaks at higher frequencies (750 cm^{-1} to 950 cm^{-1}) increases at higher nanocrystal loading, suggesting reduced connectivity of the glassy network. In particular, the highest-frequency peak, 900 cm^{-1} to 925 cm^{-1} , is assigned to terminal $\text{Nb}-\text{O}$ bonds. The ratio between these terminal bonds and bridging ones is greater in the nanocomposites (Fig. 2, Table 1, Supplementary Fig. 4), which we attribute to internal interruption of the polymerized network. Generally, the

degree of connectivity in oxide glasses can be reduced by addition of network modifiers or glass intermediates²¹. In our case, we suggest that indium (In) and tin (Sn) from the ITO nanocrystals diffuse locally into the NbO_x glass during thermal processing. There they disrupt the glassy network, inducing chain breaking and thus reducing the network connectivity. Indeed, XRD patterns of films processed at increasing temperatures (400°C – 550°C) showed broader ITO peaks (Supplementary Fig. 5b), pointing to progressive interfacial dissolution of ITO in the NbO_x matrix. The presence of a ternary InNbO_4 phase following high-temperature annealing ($>550^\circ\text{C}$) (Supplementary Fig. 5c, d) further supports the idea that In and probably Sn diffuse into the NbO_x matrix adjacent to the nanocrystals. Hence, the structure of the NbO_x glass is profoundly altered by covalent linking to ITO nanocrystals, becoming highly distorted, less interconnected and interfacially doped (Fig. 3).

The consequence of linking ITO nanocrystals within a NbO_x glass matrix is that a unique electrochromic optical switching functionality can be realized (Fig. 4). As explained below, this new optical behaviour not only reflects the individual properties of the two components, but also is quantifiably superior as a result of the reconstruction of the amorphous material.

Amorphous metal oxides are leading candidates for dynamically controlling solar radiation transmitted through ‘smart’ windows by means of their electrochromic properties. Electrochromism is a reversible change of optical transmittance in response to electrochemical charging and discharging. After more than 40 years of research in electrochromics²², spectral selectivity, that is, independent modulation of visible and near-infrared (NIR) radiation, is still considered a ‘holy grail’ for reducing the energy needed to light and thermally regulate building interiors. For instance, others have proposed combining thermochromic and electrochromic materials to modulate NIR and visible light, respectively²³, although fully integrated devices based on this concept have yet to be reported. We took a step towards this goal through our recent demonstration that charged ITO nanocrystals selectively block NIR light through a plasmonic electrochromic effect²⁴ (Fig. 4f). In contrast, amorphous transition metal oxides such as NbO_x modulate mostly visible light under electrochemical bias (Fig. 4a). Now, by linking these two materials and controlling the applied voltage, we find that the transmittance can be modulated in both spectral bands (Fig. 4a–g, Supplementary Fig. 1). The nanocrystal-in-glass coating switches progressively between three optical states: fully transparent (at 4 V versus Li), selectively NIR blocking (around 2.3 V) and broadband blocking of visible and NIR (1.5 V). Thus, solar radiation can now be dynamically modulated with spectral selectivity (Supplementary Fig. 1). The modulation of NIR light derives from the free electrons in the ITO, whose concentration reaches approximately $1.9 \times 10^{21} \text{ cm}^{-3}$ at 2.3 V (Supplementary Information and Supplementary Fig. 9). Ultimately, the modular construction of our materials will facilitate even greater modulation of NIR transmittance as new plasmonic nanocrystals are developed with a higher free-electron concentration and a broader dynamic optical range²⁵.

Remarkably, the modest visible-light modulation of amorphous NbO_x (Fig. 4a) was greatly exceeded by that of the nanocrystal-in-glass composites (Fig. 4b–d), indicating that the reconstructed NbO_x glass is far more electrochemically active. The optical contrast in the visible range monotonically increased with nanocrystal content up to $\sim 43\%$ ITO,

Table 1 | Peak-fitting analysis of Raman spectra

ITO nanocrystals (vol.%)	Specific interface area (nm^{-1})	Peak position (cm^{-1})		Peak area ratios	
		Bridging oxygens in edge-shared $[\text{NbO}_6]$	Bridging oxygens in vertex-shared $[\text{NbO}_6]$	Edge-shared/ vertex-shared	Bridging oxygens /terminal oxygens
64	0.85	651	559	6.9	5.1
52	0.69	656	574	5.9	3.7
48	0.64	664	585	2.9	6.7
0	0	675	598	1.1	8.4

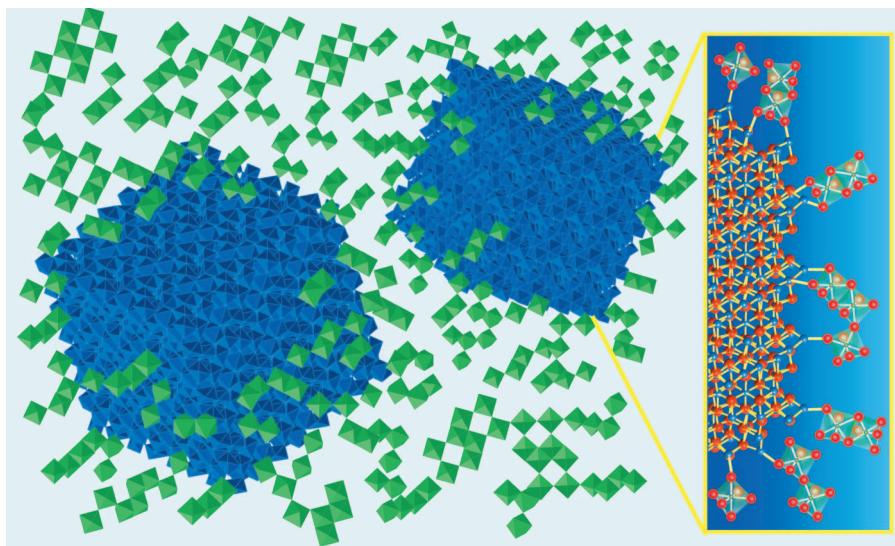


Figure 3 | ITO nanocrystals covalently linked to amorphous NbO_x . The $[\text{NbO}_6]$ octahedral units are shown in green and the ITO nanocrystals are blue.

In the inset, the niobium atoms are located at the centre of the green octahedra, oxygen atoms are red and indium atoms are blue.

reaching a value that is five times greater than that of pure NbO_x (Fig. 4h). At higher nanocrystal loading, the contrast decreased again, indicating that the enhancement is maximized when the average nanocrystal spacing is around 0.7 nm (Supplementary Information). At this high interface density (0.54 nm^{-1}), the structure of the NbO_x is rearranged throughout, effectively resulting in a new polymorph that exhibits functional characteristics distinct from those of the bulk glass, in fact resembling

more the stronger electrochromism seen in the high-temperature pseudohexagonal crystalline phase of Nb_2O_5 (known as the TT-phase) (ref. 26).

The reconstructed, nanocrystal-linked amorphous NbO_x network is also highly stable to insertion and extraction of Li ions. The electrochemical cycling stability exceeds that of either individual component; only 4% charge capacity was lost after 2,000 cycles (between 4 V and 1.5 V), whereas pure NbO_x and especially ITO nanocrystal films degraded substantially (20% and 85% capacity losses, respectively) (Supplementary Fig. 7a). In agreement with the electrochemical stability of ITO-in- NbO_x composites, the dual-band optical modulation characteristics were maintained after cycling (Supplementary Fig. 7b). We suggest that the enhancements in both optical contrast and stability are related to the more open network structure of the reconstructed glass, which facilitates ion insertion and extraction and helps to relax the stress induced by these processes. Consistent with this hypothesis, the charge capacity increases proportionally with the optical contrast (Supplementary Fig. 8).

Linking undoped In_2O_3 nanocrystals to NbO_x instead of Sn-doped In_2O_3 nanocrystals, similarly enhanced the optical contrast arising from the matrix (Fig. 4h, Supplementary Fig. 10), which eliminates the possibility that potential electrochemically active Sn species could play a part in such enhancement. Instead, this result further confirms that the optical contrast enhancement is related to a structural reconstruction of the NbO_x matrix as a consequence of the covalent linkage to the embedded nanocrystals. Very probably, similar rearrangements of the glass structure could be involved in the enhanced ionic transport properties previously reported^{13,14,27} in diverse nanocrystal-in-glass materials made by a variety of methods. Therefore, reconstructing glass

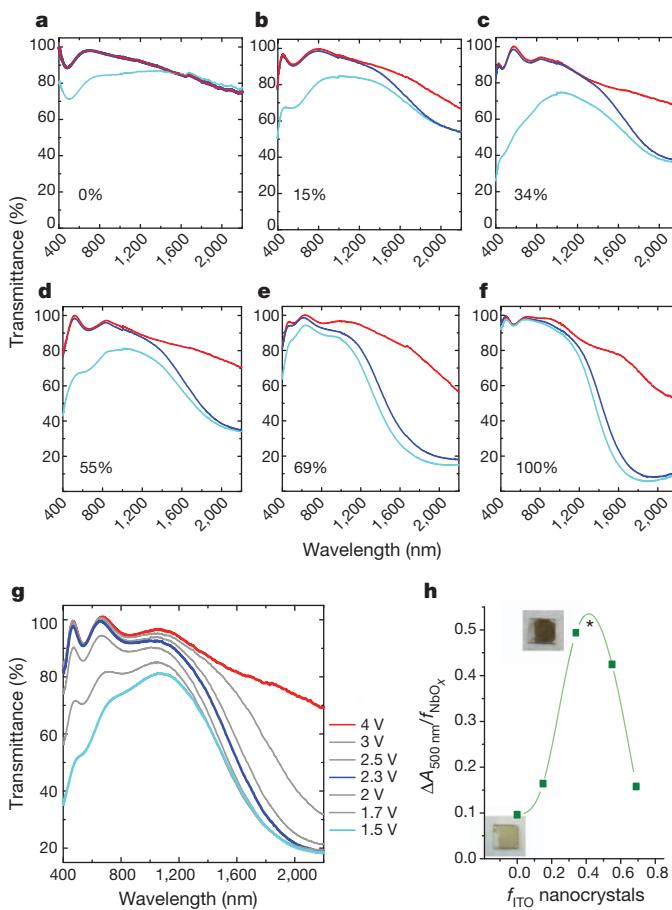


Figure 4 | Tunable dual-band solar control and optical contrast enhancement in nanocrystal-in-glass films. **a–g**, Transmittance spectra under applied electrochemical voltage (4 V in red, 2.3 V in blue and 1.5 V in cyan, versus Li/Li^+), of ITO-in- NbO_x with different ITO nanocrystal content. Pure NbO_x (a) and pure ITO nanocrystal films (f) are shown as reference. The specific interfacial area was tuned from 0 nm^{-1} to 0.92 nm^{-1} by varying the ITO nanocrystal percentage by volume between 0% and 69% at a constant size of $4.8 \text{ nm} \pm 0.8 \text{ nm}$, chosen to maximize the electrochromic response²⁴. The atomic percentage of Sn in ITO was $12.2 \pm 0.1\%$ and film thickness was $150 \pm 35 \text{ nm}$. **g**, Transmittance spectra of a typical ITO-in- NbO_x film, with intermediate voltages shown in grey. **h**, Change in absorbance (ΔA) at $\lambda = 500 \text{ nm}$ between 1.5 V and 4 V, versus volume fraction of ITO nanocrystals (f_{ITO}). ΔA is normalized by NbO_x volume fraction. The ΔA of a control sample with undoped In_2O_3 nanocrystals is also plotted (star symbol). Photographs of charged films with $f_{\text{ITO}}=0$ and $f_{\text{ITO}}=0.34$ are inset.

via nanocrystal linking could represent a general strategy for the manipulation of ion transport.

In essence, this nanocrystal-in-glass approach represents the next step in the evolution of the concept of linking building units together to construct new functional materials. Analogously, organic molecules are held together by weak, dynamic bonds to form supramolecular assemblies²⁸, metal ions are covalently bonded to organic struts to form metal–organic frameworks²⁹ and nanocrystals are linked directly to each other in mesoporous architectures³⁰. We have shown that when nanocrystals are covalently bonded to glass, the linking plays a far more active part than in these earlier constructs: the glass structure can be profoundly changed, generating new amorphous structural motifs with distinct functionality.

METHODS SUMMARY

Nanocrystal-in-glass (ITO-in-NbO_x) composite films were prepared from aqueous colloidal solutions, containing polyoxometalate clusters bonded to the nanocrystals' surface. On solution deposition (spin coating) and thermal annealing (400 °C in air), polyoxometalates condense into an amorphous oxide matrix, which remains covalently linked to the embedded nanocrystals. The volume fraction of nanocrystals in the nanocomposite films was tuned by varying the polyoxometalate molar concentration in the initial colloidal solution. XRD and scanning transmission electron microscopy were used for structural characterization of the films at different stages of the thermal process. In addition, films with different interfacial densities were analysed by Raman spectroscopy to systematically follow structural changes in the amorphous matrix induced by the presence of nanocrystals. In parallel, the electrochromic properties were studied by means of spectroelectrochemical measurements. Transmittance spectra were acquired *in situ*, under applied voltage, by placing the film as working electrode in a liquid spectroelectrochemical cell (Li foils as counter/reference electrodes and LiClO₄/propylene carbonate as electrolyte).

Full Methods and any associated references are available in the online version of the paper.

Received 14 December 2012; accepted 12 June 2013.

1. Lines, M. E. Oxide glasses for fast photonic switching—a comparative study. *J. Appl. Phys.* **69**, 6876–6884 (1991).
2. Kim, S. H. & Yoko, T. Non-linear optical properties of TeO₂-based glasses: MO_(x)–TeO₂ (M=Sc, Ti, V, Nb, Mo, Ta, and W) binary glasses. *J. Am. Ceram. Soc.* **78**, 1061–1065 (1995).
3. Nomura, K. et al. Room-temperature fabrication of transparent flexible thin-film transistors using amorphous oxide semiconductors. *Nature* **432**, 488–492 (2004).
4. Arhammar, C. et al. Unveiling the complex electronic structure of amorphous metal oxides. *Proc. Natl Acad. Sci. USA* **108**, 6355–6360 (2011).
5. Kim, Y.-H. et al. Flexible metal-oxide devices made by room-temperature photochemical activation of sol-gel films. *Nature* **489**, 128–132 (2012).
6. Idota, Y., Kubota, T., Matsufuji, A., Maekawa, Y. & Miyasaka, T. Tin-based amorphous oxide: a high-capacity lithium-ion-storage material. *Science* **276**, 1395–1397 (1997).
7. Granqvist, C. G. *Handbook of Inorganic Electrochromic Materials* (Elsevier Science, 2002).
8. Tangirala, R., Baker, J. L., Alivisatos, A. P. & Milliron, D. J. Modular inorganic nanocomposites by conversion of nanocrystal superlattices. *Angew. Chem. Int. Ed.* **49**, 2878–2882 (2010).
9. Llordes, A. et al. Polyoxometalates and colloidal nanocrystals as building blocks for metal oxide nanocomposite films. *J. Mater. Chem.* **21**, 11631–11638 (2011).
10. Rosenblanz, A. et al. Bulk glasses and ultrahard nanoceramics based on alumina and rare-earth oxides. *Nature* **430**, 761–764 (2004).
11. Falcão-Filho, E. L. et al. Third-order optical nonlinearity of a transparent glass ceramic containing sodium niobate nanocrystals. *Phys. Rev. B* **69**, 134204 (2004).
12. Mattarelli, M., Gasperi, G., Montagna, M. & Verrocchio, P. Transparency and long-ranged fluctuations: the case of glass ceramics. *Phys. Rev. B* **82**, 094204 (2010).

13. Schirmisen, A. et al. Fast interfacial ionic conduction in nanostructured glass ceramics. *Phys. Rev. Lett.* **98**, 225901 (2007).
14. Zhou, H. S., Li, D. L., Hibino, M. & Honma, I. A self-ordered, crystalline-glass, mesoporous nanocomposite for use as a lithium-based storage device with both high power and high energy densities. *Angew. Chem. Int. Ed.* **44**, 797–802 (2005).
15. Dong, W. et al. Controllable and repeatable synthesis of thermally stable anatase nanocrystal-silica composites with highly ordered hexagonal mesostructures. *J. Am. Chem. Soc.* **129**, 13894–13904 (2007).
16. Sakamoto, A. & Yamamoto, S. Glass-ceramics: engineering principles and applications. *Int. J. Appl. Glass Sci.* **1**, 237–247 (2010).
17. Dong, A. et al. A generalized ligand-exchange strategy enabling sequential surface functionalization of colloidal nanocrystals. *J. Am. Chem. Soc.* **133**, 998–1006 (2011).
18. Rosen, E. L. et al. Exceptionally mild reactive stripping of native ligands from nanocrystal surfaces by using Meerwein's salt. *Angew. Chem. Int. Ed.* **51**, 684–689 (2012).
19. McConnell, A. A., Anderson, J. S. & Rao, C. N. R. Raman-spectra of niobium oxides. *Spectrochim. Acta* **32**, 1067–1076 (1976).
20. Jehng, J. M. & Wachs, I. E. Structural chemistry and Raman-spectra of niobium oxides. *Chem. Mater.* **3**, 100–107 (1991).
21. Shelby, J. E. *Introduction to Glass Science and Technology* Ch. 2, 5 (Royal Society of Chemistry, 2005).
22. Monk, P., Mortimer, R. & Rosseinsky, D. *Electrochromism and Electrochromic Devices* Ch. 2 (Cambridge Univ. Press, 2007).
23. Li, S.-Y., Niklasson, G. A. & Granqvist, C. G. Nanothermochromics: calculations for VO₂ nanoparticles in dielectric hosts show much improved luminous transmittance and solar energy transmittance modulation. *J. Appl. Phys.* **108**, 063525 (2010).
24. Garcia, G. et al. Dynamically modulating the surface plasmon resonance of doped semiconductor nanocrystals. *Nano Lett.* **11**, 4415–4420 (2011).
25. Garcia, G. et al. Near-infrared spectrally selective plasmonic electrochromic thin films. *Adv. Opt. Mater.* **1**, 215–220 (2013).
26. Rosario, A. V. & Pereira, E. C. Influence of the crystallinity on the Li⁺ intercalation process in Nb₂O₅ films. *J. Solid State Electrochem.* **9**, 665–673 (2005).
27. Wang, R. Y., Tangirala, R., Raoux, S., Jordan-Sweet, J. L. & Milliron, D. J. Ionic and electronic transport in Ag₂S nanocrystal–GeS₂ matrix composites with size-controlled Ag₂S nanocrystals. *Adv. Mater.* **24**, 99–103 (2012).
28. Lehn, J. M. Supramolecular chemistry—receptors, catalysis, and carriers. *Science* **227**, 849–856 (1985).
29. Li, H., Eddaoudi, M., O'Keeffe, M. & Yaghi, O. M. Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature* **402**, 276–279 (1999).
30. Buonsanti, R. et al. Assembly of ligand-stripped nanocrystals into precisely controlled mesoporous architectures. *Nano Lett.* **12**, 3872–3877 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Raoux and J. L. Jordan-Sweet as well as S. Mannsfeld and M. Toney for assistance in synchrotron XRD measurements at the National Synchrotron Light Source (Brookhaven National Laboratory) and Stanford Synchrotron Radiation Lightsource (SSRL); and R. Zuckermann, P. J. Schuck, R. J. Mendelsberg, and especially M. Salmeron and O. Yaghi for critical reading of the manuscript. This work was performed at the Molecular Foundry, Lawrence Berkeley National Laboratory, and was supported by the Office of Science, Office of Basic Energy Sciences, of the US Department of Energy (DOE) under contract number DE-AC02–05CH11231. D.J.M. and G.G. were supported by a DOE Early Career Research Program grant under the same contract, and J.G. was supported by Consejo Superior de Investigaciones Científicas, CSIC, JAE. Scanning transmission electron microscopy images were taken at Oak Ridge National Laboratory (ORNL), supported by DOE-BES, Materials Sciences and Engineering Division, and by ORNL's Shared Research Equipment (ShaRE) User Program, which is also sponsored by DOE-BES. XRD data shown in the manuscript was acquired at SSRL, beamline 11-3.

Author Contributions A.L. synthesized the materials, carried out the experiments and analysed the data, with assistance from G.G. for the electrochemical characterization. J.G. carried out scanning transmission electron microscopy imaging. A.L. and D.J.M. were responsible for experimental design and wrote the manuscript, which incorporates critical input from all authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.J.M. (dmilliron@lbl.gov).

METHODS

Nanocrystal-in-glass (ITO-in-NbO_x) composite films were prepared from aqueous colloidal solutions containing polyoxometalate clusters bonded to the nanocrystals' surface. On solution deposition (spin coating) and thermal annealing (400 °C in air), polyoxometalates condense into an amorphous oxide matrix, which remains covalently linked to the embedded nanocrystals. The volume fraction of nanocrystals in the nanocomposite films was tuned by varying the polyoxometalate molar concentration in the initial colloidal solution. XRD and scanning transmission electron microscopy were used for structural characterization of the films at different stages of the thermal process. In addition, films with different nanocrystal interfacial areas were analysed by Raman spectroscopy to systematically follow structural changes in the amorphous matrix induced by the presence of nanocrystals. In parallel, the electrochromic properties were studied by means of spectro-electrochemical measurements. Transmittance spectra were acquired *in situ*, under applied voltage, by placing the film as working electrode in a liquid spectro-electrochemical cell (Li foils as counter/reference electrodes and LiClO₄/propylene carbonate as electrolyte).

Synthesis of POMs and ITO nanocrystals. We used two different POMs; the hexaniobates [Nb₆O₁₉]⁸⁻ and decaniobates [Nb₁₀O₂₈]⁶⁻ were both prepared using a solution precipitation method with tetramethylammonium, [N(CH₃)₄]⁺, as counterion. We recently reported the synthesis and detailed characterization of these POMs⁹. Briefly, to synthesize decaniobate clusters, 3.19 g (10 mmol) of niobium ethoxide (Nb(OEt)₅, 99.99%, Alfa Aesar) were added to a flask containing 25 ml ethanol. Then, 1.81 g (10 mmol) tetramethylammonium hydroxide (N(CH₃)₄OH·5H₂O, 97%, Aldrich) were added to the flask and the reaction mixture was refluxed overnight, then cooled to room temperature and centrifuged. The resulting solid product was washed with H₂O and precipitated with acetone (three times). After drying overnight under vacuum at room temperature, a crystalline white powder was collected (47% reaction yield). XRD of the resultant powder as well as thermogravimetry coupled to mass spectrometry confirmed the expected chemical formula: [N(CH₃)₄]₆Nb₁₀O₂₈·6H₂O (ref. 9). The hexaniobates, [Nb₆O₁₉]⁸⁻, were synthesized by dissociating decaniobate clusters at pH 14. The reaction was performed at room temperature by mixing 0.21 mmol [N(CH₃)₄]₆Nb₁₀O₂₈·6H₂O and 1.48 mmol N(CH₃)₄OH·5H₂O in a flask containing 10 ml EtOH/H₂O (1:1). After 2 h stirring, the solid product, with the expected chemical formula—[N(CH₃)₄]₈Nb₆O₁₉·15H₂O—was precipitated with acetone and dried overnight under vacuum. Given that the pH stability of hexaniobates is within the range of 12.5–14, we avoided further washing, which would induce its condensation to decaniobate clusters (stable at neutral pH)³¹.

ITO nanocrystals, with sizes ranging from 5 nm to 15 nm and atomic percentage of Sn doping from 4 at.% to 12 at.%, were synthesized by following established colloidal methods^{24,32}. The size of the nanocrystals was determined by collecting statistics in bright-field transmission electron microscope images, and atomic doping concentration was assessed by inductively coupled plasma atomic emission spectroscopy. The ITO-in-NbO_x composite films that we used for electrochromic properties testing and Raman characterization contained ITO nanocrystals with Sn at. % of 12.2 ± 0.1 and diameter 4.8 nm ± 0.1 nm.

Linking process for POM-stabilized nanocrystals. As-synthesized ITO nanocrystals were stripped of their coordinated organic ligands and their naked, metal-adatom-rich surfaces^{17,18} were further functionalized with POMs. The chemical link between POMs and the nanocrystals is made by bonding nucleophilic POM oxygens to In/Sn adatoms in ITO.

In the ligand-stripping process, a two-phase mixture was prepared in a vial by combining equal volumes of *N,N*-dimethylformamide (DMF) and organic-capped ITO (~60 mg ml⁻¹) dispersed in hexane. Then, ~7 mg nitrosonium tetrafluoroborate (NOBF₄, 97%, Acros) were added and the resulting mixture was either vigorously stirred or sonicated. After 5–10 min, ITO nanocrystals were quantitatively transferred to the polar DMF phase as a result of their organic-free surface, now weakly coordinated to the solvent (DMF) molecules. These naked nanocrystals were precipitated with toluene and the supernatant removed by centrifugation. After further purification with toluene and DMF (three washings), the nanocrystals were dispersed in DMF (final concentration, 60 mg ml⁻¹). Next, 1 ml of an aqueous solution of hexaniobate clusters, [N(CH₃)₄]₈Nb₆O₁₉ (concentration varies depending on the targeted final volume fraction, typically 200 mg ml⁻¹) were added to 1 ml DMF ITO dispersion, which instantly turned opaque. After vigorously stirring (30 min), the slurry becomes clear as a result of POMs bonding to the naked ITO nanocrystals, which now become perfectly soluble in water (Fig. 1a). The colloidal solution was washed three times with DMF/H₂O to ensure complete removal of DMF molecules. The resultant POM-ITO pellet was redispersed in ultrapure water. Note that hexaniobates were found to be more effective than

decaniobates in coordinating the bare nanocrystal surfaces to form stable aqueous dispersions. We ascribe this behaviour to the higher charge density of hexaniobates, which facilitates the surface coverage and extent of linking to the nanocrystal.

Addition of excess POM (that is, decaniobates) to the POM-ITO solution allowed us to vary the volume fraction of NbO_x glass in the final ITO-in-NbO_x film. Aqueous POM-ITO solutions are perfectly stable and no ageing has been observed after two years following preparation. Ethanol was added as co-solvent to increase the viscosity of the solution and thus improve the quality of the deposited film.

Preparation of ITO-in-NbO_x nanocrystal-in-glass films. POM-ITO colloidal solutions, in water/ethanol (1:1), were deposited by spin coating onto a substrate (such as Si or ITO-coated glass) which had been previously cleaned and treated with ultraviolet-ozone to ensure a hydrophilic surface. As-deposited POM-ITO films were then annealed at 400 °C for 20 min, in air, using a rapid thermal annealing furnace at 10 °C s⁻¹.

XRD of POM-ITO and ITO-in-NbO_x. The atomic structure of as-deposited POM-ITO (Supplementary Fig. 3) and annealed ITO-in-NbO_x films (Fig. 2b–d, Supplementary Fig. 5) was assessed by XRD with synchrotron radiation ($\lambda = 0.974 \text{ \AA}$), at the Stanford Synchrotron Radiation Lightsource (SSRL), beamline 11-3. Acquisitions were performed in grazing incident mode (incidence angle of 0.12°) and the scattered X-rays collected with an area detector.

Transmission electron microscopy of POM-ITO and ITO-in-NbO_x. Transmission electron microscope images of the POM-ITO colloidal solutions (Supplementary Fig. 2) were acquired on a JEOL-2100 with a LaB₆ filament at 200 kV and equipped with a Gatan camera. Samples for transmission electron microscope analysis were prepared by drying a drop of POM-ITO aqueous solution on the surface of a carbon-coated copper grid, which had been previously treated with ultraviolet-ozone.

Cross-sectional scanning transmission electron microscopy images of ITO-in-NbO_x films (Fig. 1e, f) were carried out in a Nion UltraSTEM operated at 100 kV and equipped with a Nion aberration corrector. A high-angle annular dark field (HAADF) detector allowed us to record incoherent Z-contrast images, in which the intensity of atom columns directly reflects their mean square atomic number (Z). Specimens for scanning transmission electron microscopy were prepared by conventional methods, that is, grinding, dimpling and Ar ion milling.

Scanning electron microscopy and energy dispersive spectroscopy. Scanning electron microscope imaging (Supplementary Fig. 6) was carried out on a Zeiss Gemini Ultra-55 microscope, using beam energies of 5–10 kV and an In-Lens detector. Film thickness was determined by cross-sectional imaging. An inbuilt energy dispersive spectroscopy detector was used for elemental analysis. The volume fraction of each component in the final nanocrystal-in-glass films was estimated from the measured metal atomic ratios and their bulk densities.

Raman spectroscopy of ITO-in-NbO_x films. The bonding arrangement of the amorphous NbO_x matrix was studied by Raman spectroscopy (Fig. 2), using a Horiba LabRAM Aramis instrument with confocal aperture. All Raman spectra were obtained with a ×100 microscope objective (numerical aperture of 0.90), 17.7 mW of laser power at an excitation wavelength of 532 nm, and an acquisition time of 300 s. The spectral resolution was 1.5 cm⁻¹ and the laser spot size was ~721 nm. Several spots were examined on each sample and showed similar spectra. Raman peaks from the substrate (ITO-sputtered glass slides) appeared in a different spectral region than NbO_x, allowing accurate data analysis and interpretation. Peak-fitting analysis was carried out in OriginPro software.

Optical switching and electrochemical measurements of ITO-in-NbO_x films. Optical switching properties at different voltages (Fig. 4) were studied in a home-built spectroelectrochemical cell, in which the ITO-in-NbO_x film was placed as the working electrode and *in situ* transmission visible–NIR spectra were recorded as functions of the applied potential (1.5 V–4 V versus Li/Li⁺). The spectroelectrochemical cell had a three-electrode configuration, with Li foils as counter and reference electrodes, the electrolyte being 0.1 M LiClO₄ in anhydrous propylene carbonate.

The electrochemical stability was evaluated by cyclic voltammetry, in which the voltages were swept between the potential limits at 30 mV s⁻¹ 2,000 times. Spectroelectrochemical and cyclic voltammetry measurements were performed in an argon glove box with a ASD Quality Spec Pro visible–NIR spectrometer and a Bio-logic VSP potentiostat. ITO-in-NbO_x films were deposited on conducting glass substrates (that is, ITO-sputtered glass slides) to ensure good in-plane conductivity.

31. Villa, E. M. et al. Reaction dynamics of the decaniobate ion [H_xNb₁₀O₂₈]^{(6-x)-} in water. *Angew. Chem. Int. Edn* **47**, 4844–4846 (2008).
32. Choi, S. I., Nam, K. M., Park, B. K., Seo, W. S. & Park, J. T. Preparation and optical properties of colloidal, monodisperse, and highly crystalline ITO nanoparticles. *Chem. Mater.* **20**, 2609–2611 (2008).

No increase in global temperature variability despite changing regional patterns

Chris Huntingford¹, Philip D. Jones^{2,3}, Valerie N. Livina^{4,5}, Timothy M. Lenton⁶ & Peter M. Cox⁷

Evidence from Greenland ice cores shows that year-to-year temperature variability was probably higher in some past cold periods¹, but there is considerable interest in determining whether global warming is increasing climate variability at present^{2–6}. This interest is motivated by an understanding that increased variability and resulting extreme weather conditions may be more difficult for society to adapt to than altered mean conditions³. So far, however, in spite of suggestions of increased variability², there is considerable uncertainty as to whether it is occurring⁷. Here we show that although fluctuations in annual temperature have indeed shown substantial geographical variation over the past few decades², the time-evolving standard deviation of globally averaged temperature anomalies has been stable. A feature of the changes has been a tendency for many regions of low variability to experience increases, which might contribute to the perception of increased climate volatility. The normalization of temperature anomalies² creates the impression of larger relative overall increases, but our use of absolute values, which we argue is a more appropriate approach, reveals little change. Regionally, greater year-to-year changes recently occurred in much of North America and Europe. Many climate models predict that total variability will ultimately decrease under high greenhouse gas concentrations, possibly associated with reductions in sea-ice cover. Our findings contradict the view that a warming world will automatically be one of more overall climatic variation.

A recent analysis² of the NASA Goddard Institute for Space Studies gridded temperature data set⁸ suggests that seasonal mean temperatures are starting to show increased variability, with a particularly large difference between the 1970s and the 1980s. The analysis first calculates for each geographical position the standard deviation of temperature for period 1951–1980. It then divides each local time series of anomalies away from a mean (also calculated for 1951–1980) by those standard deviation values, and hence provides a normalization of year-to-year variations on to a single distribution for all points across the globe. When the distribution is presented for different decades, it is shown to widen, particularly when comparing the periods before and after 1980, indicating major changes in climate variability. Some features of climate shifts around that time have been linked to changes in the Pacific Decadal Oscillation^{9,10}. In this work, we perform an analysis similar to that in ref. 2, using the European Centre for Medium-Range Weather Forecasts (ECMWF) ERA-40¹¹ data set that starts in 1958, and generate analogous distribution curves (Fig. 1a). The change in standard deviation between the distribution for 1958–1970 and that for 1991–2001 (Fig. 1a) is an increase of 43% (Supplementary Table 1, first row). There are seasonal and hemispheric variations in these distributions (Supplementary Fig. 1 and Supplementary Table 1), although the common feature of generally increasing variance remains present in all cases. All statistical details are described in Methods.

One contribution to the increase over time of the standard deviations of the distributions, when using the algorithm leading to Fig. 1a,

is geographically diverse rates of mean warming. To account for this, we reconsider the curves of Fig. 1a and, for each of the four periods, instead calculate anomalies relative to the average of each period (as opposed to always relative to 1958–1980). Division is again made by local standard deviations calculated for before 1980, but now for these new anomalies, giving the revised curves in Fig. 1b. Although this reduces the changes to the distributions between periods, there remains an increase of 22% in standard deviation between 1958–1970 and 1991–2001 (Supplementary Table 2, first row).

To understand further, we instead derive at each ERA-40 data set grid box the yearly temperature anomaly (detrending with an 11-year local running mean) and then calculate a time-evolving, area-weighted, global, annual standard deviation from these. This time series of total variability is the black curve in Fig. 1c. Although there is evidence of some increase around 1980, this is an order of magnitude less than would be implied by ref. 2 and our Fig. 1a or Fig. 1b. The mean of the black curve in Fig. 1c before and after 1980 shows only a 2% change. We then normalize the value at each point by the local standard deviation calculated up to 1980, and re-derive the yearly, global standard deviation values based on this. We observe these values to increase more (Fig. 1c, red curve): the mean after 1980 is 16% greater than the mean before 1980, which is similar to the change calculated from Fig. 1b. The green curve in Fig. 1c is the ratio of the red and black curves. In Fig. 1d, we plot, for local standard deviations of the anomalies from detrending, calculated before and after 1980, the number of grid boxes in which the standard deviation subsequently increases (red curve) or decreases (green curve) after 1980. The curves are not similar and there is a tendency for low-variance regions to have a higher chance of becoming more variable. We offer this as an explanation why, overall, normalized variability has increased more. In terms of relative magnitude, an increase in variance for regions of low standard deviation will outweigh a similar decrease for a region of high variation. Absolute global standard deviation for the ERA-Interim¹² period (1984–2006) is also plotted (Fig. 1c, cyan curve; overlapping the black curve for 1984–1996), and, rather than increasing, is actually found to be decreasing in the most recent years.

Nevertheless, changes to regional variability have occurred. Supplementary Table 2 (and Supplementary Fig. 2) suggests that the largest changes in standard deviation, by season and hemisphere, are in the Northern Hemisphere spring and summer and the Southern Hemisphere autumn and summer. There is already some evidence that increased variability for Europe, imposed additionally on general warming, is increasing the frequency of European summer heatwaves¹³. Supplementary Fig. 3 shows the average zonal changes to standard deviation before and after 1980, with prominent increases over mid latitudes for both hemispheres. We note observations from the hydrological cycle indicating that storm tracks have moved polewards in recent decades^{14–16}, along with Hadley cell expansion^{17,18}. To confirm more directly than by use of gridded re-analysis data that major temperature

¹Centre for Ecology and Hydrology, Wallingford OX10 8BB, UK. ²Climatic Research Unit, School of Environmental Sciences, Faculty of Science, University of East Anglia, Norwich NR4 7TJ, UK. ³Center of Excellence for Climate Change Research, Department of Meteorology, King Abdulaziz University, Jeddah 21589, Saudi Arabia. ⁴School of Environmental Sciences, Faculty of Science, University of East Anglia, Norwich NR4 7TJ, UK. ⁵National Physical Laboratory, Hampton Road, Teddington TW11 0LW, UK. ⁶College of Life and Environmental Sciences, University of Exeter, Exeter EX4 4PS, UK. ⁷College of Engineering, Mathematics and Physical Sciences, University of Exeter, Exeter EX4 4QF, UK.

variability changes are occurring, we analyse weather station data¹⁹ that contributes to the University of East Anglia Climatic Research Unit (CRU) climatology. We detrend each station data set of calculated yearly mean temperatures and derive a time-evolving series of standard deviations. To illustrate behaviour before and after 1980, we scale such that

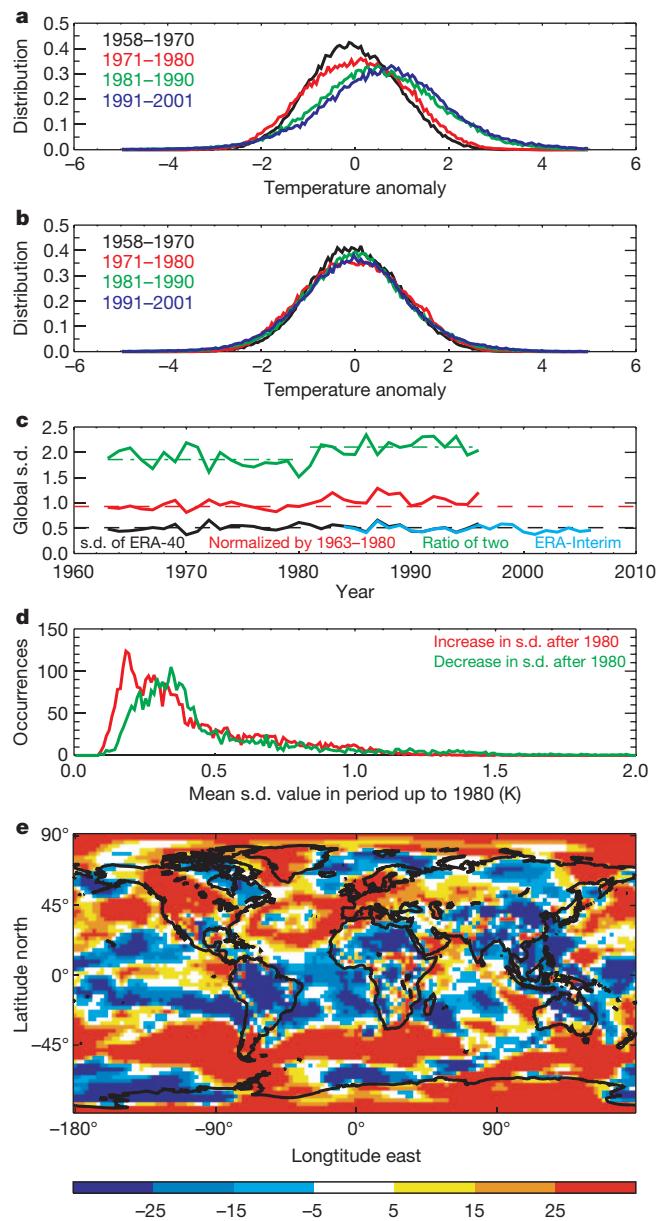


Figure 1 | Measured changes in temperature variability. **a**, Normalized probability density functions for four periods of anomalies in the ECMWF ERA-40 data set. Anomalies are calculated relative to 1958–1980 mean values and then divided by their local standard deviation for that period. **b**, Anomalies calculated relative to the mean for each of the four periods and then divided by the standard deviation of these anomalies for 1958–1980. **c**, Global, yearly spatial average of standard deviations of anomalies from local, 11-yr detrending. Black curve, ERA-40; cyan, ERA-Interim (K). The only difference between the red curve and the black is that, for the former, the anomalies are divided by their local standard deviation for 1963–1980. The green curve is the ratio of the red and black curves. The black and red dashed lines are the average values of the correspondingly coloured curves to 1980. The green dash-dot lines are mean values of the green curve before and after 1980. **d**, Grid-box (area-weighted, detrended data) occurrences, for bins of size 0.01 K, showing increases (red) and decreases (green) in standard deviation in 1981–1996 compared with 1963–1980. **e**, Percentage change in standard deviation of ERA-40 data, again comparing the periods after and before 1980: yellow, orange and red indicate more variability after 1980.

each series has mean of unity between 1975 and 1985. The spread of curves in Supplementary Fig. 4a is large before and after 1980, confirming significant change. In Supplementary Fig. 4b, c, we show only stations in mid Europe and North America; here variability predominantly increased during the 1970s and 1980s, but less so in more recent years.

Figure 1e shows geographical change in standard deviation based on the ERA-40 data set. Each grid point is the percentage change in standard deviation in detrended annual temperature, comparing again the periods before and after 1980. Of particular interest are the very large changes in variability: in many places there are changes of more than 25%, both increases after 1980 (red) and decreases (dark blue). North America and Europe are once more seen to have generally experienced increases in standard deviation, although the magnitude of changes is seasonally dependent (Supplementary Fig. 5; for much of Europe the largest increases occur during December–May). Of note is the strong spatial structure of the map in Fig. 1e, which has large coherent regions of change.

Hovmöller plots provide time-evolving spatial (zonal) information. Figure 2 complements the map in Fig. 1e and shows, for each year and 10° latitude band, standard deviation as a percentage of its value for 1968–1980, calculated for both the ERA-40 period and the ERA-Interim period. The increased mid-latitude standard deviation values are apparent, along with a general tropical decrease. There is again a suggestion, towards the end of the ERA-Interim data, that the zones of increased standard deviation are starting to subside, especially after 1995.

To assess whether the observed changes in variability are transitory, we analyse projections made using general circulation models. Figure 3a shows the global, time-evolving standard deviation from one such model, HadGEM2-ES²⁰, using the same method as that giving the black and cyan curves in Fig. 1c for ECMWF data. For this model, the Coupled Model Intercomparison Project Phase 5 (CMIP5) database²¹ has four ensemble members (all shown in Fig. 3a) driven first with all historical climatic forcings and then by the atmospheric greenhouse gas changes prescribed in representative concentration pathway RCP8.5²². The mean of the 11-yr moving averages of each ensemble member is the thick black curve. This predicts major ongoing reductions in variability during the twenty-first century, and suggests that there was a small increase in variability during the 1980s and 1990s. The thick red curve is the mean of the global standard deviation for ERA-40/ERA-Interim (moving average of black and cyan values in Fig. 1c), and shows that global variability first increased slightly and then decreased.

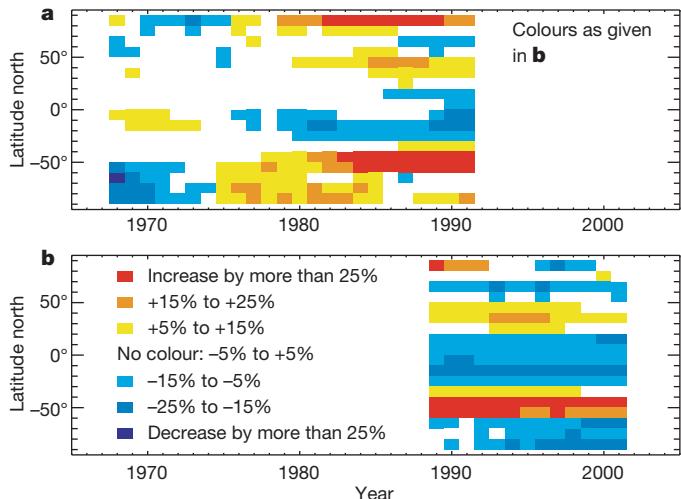


Figure 2 | Zonal changes in measured temperature variability. Hovmöller plots of time-evolving standard deviation in 10° latitudinal zonal bands. **a**, ERA-40 data set; **b**, ERA-Interim data set. All data shown as percentage change relative to zonal averages for 1968–1980.

Figure 3b–e shows the percentage changes in yearly zonal standard deviation for the four HadGEM2-ES ensembles with historical forcings and RCP8.5, compared with their mean to 1980. Eventual variance reduction occurs across most latitudinal bands, and all changes have spatial structure and some evidence of temporal persistence beyond the 11 yr implicit in the moving averages (Methods), although regional timings differ between simulations. There are predictions of transitory

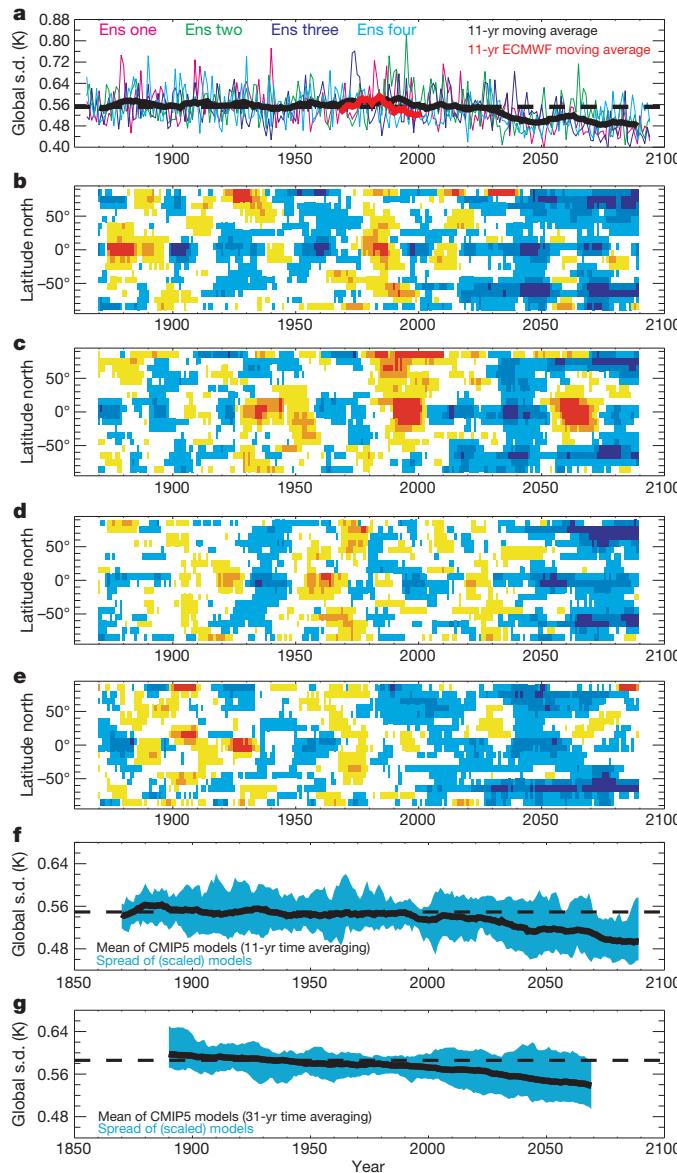


Figure 3 | Projected changes in temperature variability under the RCP8.5 scenario. a, Global, spatially averaged standard deviation in yearly temperature anomalies (from 11-yr detrending) for four HadGEM2-ES simulations, driven with all historical forcings and then the RCP8.5 scenario. The mean of the 11-yr moving averages of the ensemble members is the thick black curve; dashed line shows the average of the black curve up to 1980. The thick red curve is the 11-yr moving average of ERA-40 and ERA-Interim global standard deviations, and for comparison is scaled to have the same average value as the model ensemble average before 1980 (black dashed line). b–e, Percentage changes in zonal standard deviation relative to average before 1980, for the four simulations shown in a. The colour scale is as in Fig. 2. f, Thick black curve as in a, but averaged across CMIP5 models (for individual models, see Supplementary Fig. 6). Then each model is scaled such that its pre-1980 average is the same as the all-model mean (dashed line), which allows comparison of relative changes (turquoise model spread). g, As in f, except local anomalies use 31-yr detrending, and data plotted are 31-yr moving averages of global standard deviations.

occurrences of increased variability (orange/red patches) in recent decades, but the timings, positions and magnitudes of these variations vary markedly between ensemble members.

Figure 3f shows the mean, time-evolving, global standard deviation for all 17 climate models with historical-plus-RCP8.5 simulations in the CMIP5 database²¹ (thick black curve). Individual model standard deviation curves are shown in Supplementary Fig. 6, and approximately two-thirds show an eventual definite decrease in average variability. Scaling each model's time-evolving standard deviation curve to have the same average value (same as dashed line) before 1980 allows relative change to be seen more clearly. The bounds of these scaled curves also highlight this cross-model decrease towards the end of the twenty-first century (turquoise area). These multimodel findings are robust to changes in the timescale selected for both detrending by local means and subsequent calculation of global, moving averages of standard deviation. Figure 3g and Supplementary Fig. 7 have features in common with Fig. 3f and Supplementary Fig. 6 when instead using 31-yr timescales.

One possible reason for the reduction in global temperature variability is the decrease in Arctic sea-ice cover in these model projections. Sea-ice extent strongly affects multiple features of the global climate system^{23–25}. Changes in it have been argued to amplify, in particular, Northern Hemisphere temperature anomalies through sea-ice albedo feedbacks^{26,27}. Reductions in sea-ice thickness and age have reduced the timescale of changes in sea-ice extent, acting to increase temporarily the amplification of interannual temperature anomalies for mid latitudes²⁷. Ultimately, however, major loss of summer sea ice in extreme global warming scenarios will lead to the removal of a possible enhancement of interannual temperature anomalies. In Fig. 4, we plot, for the CMIP5 models, our global standard deviation values against sea-ice extent. This shows a general decrease in standard deviation for less sea ice. When presented as individual plots for each model (Supplementary Fig. 8), some models show an almost linear correspondence between the two quantities, potentially consistent with the hypothesis that the loss of sea ice removes an important amplifier of interannual temperature variability²⁷. However, these figures do not

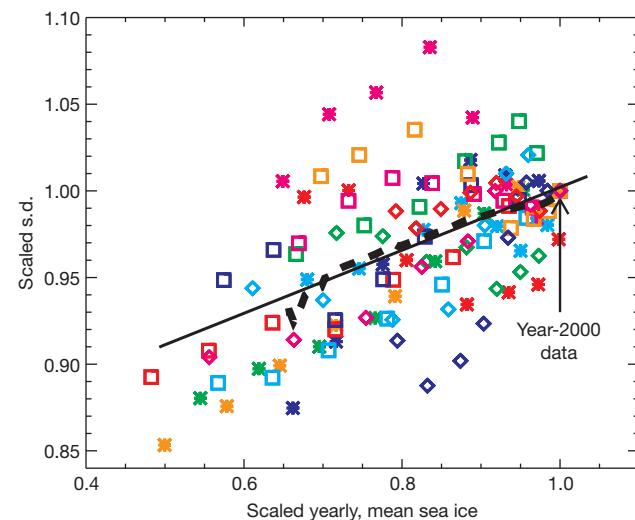


Figure 4 | Projected changes in temperature variability against changes in sea-ice cover. Thirty-one-year moving averages of CMIP5 estimates of global, annual sea-ice cover, plotted against the 31-yr moving averages of global temperature standard deviation. Non-black colours and symbols represent the various models, shown every decade after 2000 (for individual model values, see Supplementary Figs 7 and 8). For comparison, each model's predictions of standard deviation and sea ice are scaled to unity for the year 2000 (arrow). The dashed black line is the average across scaled models, covering change in sea ice over the range that most models span. The straight black line is linear fit to the dashed curve.

demonstrate a causal link, which would require additional climate change projections excluding sea-ice feedbacks.

We identify major changes in temperature fluctuations in ECMWF gridded data and CRU data from individual stations. Significant and coherent spatial shifts are detected in regions of differing temperature variability, but we find much less evidence of changes when the data are averaged globally². Increases in variability are especially evident during the 1980s and 1990s in the densely monitored regions of Europe and North America. However, even in these regions there is observational evidence that interannual temperature variability is now decreasing. Climate models strongly suggest that this might be the beginning of a longer-term decline in year-to-year temperature variability in response to increases in atmospheric greenhouse gas concentrations.

Changes in climate variability are arguably more important for society than changes in mean climate such as overall warming, especially if they relate to altered extremes^{3,28,29}. On the basis of the results presented here, we believe that greater emphasis now needs to be placed on analysing changes in climate variability in the context of anthropogenic climate change, so as to inform more effective adaptation strategies.

METHODS SUMMARY

Our data are based on yearly or seasonal averages of atmospheric temperature values 2 m above the surface, and from the ECMWF ERA-40¹¹ and ERA-Interim¹² products. These are six-hourly, globally complete data sets, and in combination cover the period from 1958 to the present. They constitute a re-analysis product, derived by assimilating meteorological measurements into a weather forecast model to retrospectively provide high-frequency, gridded estimates of evolving weather states.

For future climate projections, we use models from the CMIP5 family of climate simulations²¹. Of these models, 17 are available driven first with historical forcings (all forcings) and then RCP8.5 atmospheric gas concentration changes²².

To calculate local, zonal or global average standard deviations (Figs 1c–e, 2, 3 and 4), for each grid box, anomalies are first calculated by detrending local temperature values with a running mean (calculated over 11 yr except in Figs 3g and 4, which use 31-yr running means). Zonal or global spatial standard deviations of these anomalies are calculated with area weighting for each grid box. ERA-40 and ERA-Interim cover the years 1958 to 2011 inclusive, so with local, 11-yr detrending, global average standard deviation values (see, for example, Fig. 1c) will correspond to the remaining years of 1963–2006. When 11-yr moving averages of these time-evolving zonal or global standard deviations of the de-trended anomalies are calculated (see, for example, Fig. 2 or the thick red curve in Fig. 3a), this corresponds to a remaining period of 1968–2001. Similarly, when climate-model local anomalies are derived from 31-yr detrending and then 31-yr moving averages are calculated on the basis of the globally averaged standard deviation of these anomalies, this gives a time series for the years 1890–2070 for simulations covering 1860–2100 (Fig. 3g).

Full Methods and any associated references are available in the online version of the paper.

Received 9 January; accepted 17 May 2013.

Published online 24 July 2013.

1. Steffensen, J. P. et al. High-resolution Greenland Ice Core data show abrupt climate change happens in few years. *Science* **321**, 680–684 (2008).
2. Hansen, J., Sato, M. & Ruedy, R. Perception of climate change. *Proc. Natl Acad. Sci. USA* **109**, E2415–E2423 (2012).
3. Rahmstorf, S. & Coumou, D. Increase of extreme events in a warming world. *Proc. Natl Acad. Sci. USA* **108**, 17905–17909 (2011).
4. Stott, P. A., Stone, D. A. & Allen, M. R. Human contribution to the European heatwave of 2003. *Nature* **432**, 610–614 (2004).

5. Field, C. B. et al. (eds) *Managing the Risks of Extreme Events and Disasters to Advance Climate Change Adaptation* (Cambridge Univ. Press, 2012).
6. Coumou, D. & Rahmstorf, S. A decade of weather extremes. *Nature Clim. Change* **2**, 491–496 (2012).
7. Rhines, A. & Huybers, P. Frequent summer temperature extremes reflect changes in the mean, not the variance. *Proc. Natl Acad. Sci. USA* **110**, E546 (2013).
8. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).
9. Trenberth, K. E. Recent observed interdecadal climate changes in the northern-hemisphere. *Bull. Am. Meteorol. Soc.* **71**, 988–993 (1990).
10. Hartmann, B. & Wendler, G. The significance of the 1976 Pacific climate shift in the climatology of Alaska. *J. Clim.* **18**, 4824–4839 (2005).
11. Uppala, S. M. et al. The ERA-40 re-analysis. *Q. J. R. Meteorol. Soc.* **131**, 2961–3012 (2005).
12. Dee, D. P. et al. The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
13. Schär, C. et al. The role of increasing temperature variability in European summer heatwaves. *Nature* **427**, 332–336 (2004).
14. Seidel, D. J., Fu, Q., Randel, W. J. & Reichler, T. J. Widening of the tropical belt in a changing climate. *Nature Geosci.* **1**, 21–24 (2008).
15. Bender, F. A. M., Ramanathan, V. & Tselioudis, G. Changes in extratropical storm track cloudiness 1983–2008: observational support for a poleward shift. *Clim. Dyn.* **38**, 2037–2053 (2012).
16. Zhou, Y. P., Xu, K. M., Sud, Y. C. & Betts, A. K. Recent trends of the tropical hydrological cycle inferred from Global Precipitation Climatology Project and International Satellite Cloud Climatology Project data. *J. Geophys. Res.* **116**, D09101 (2011).
17. Johanson, C. M. & Fu, Q. Hadley cell widening: model simulations versus observations. *J. Clim.* **22**, 2713–2725 (2009).
18. Lu, J., Deser, C. & Reichler, T. Cause of the widening of the tropical belt since 1958. *Geophys. Res. Lett.* **36**, L03803 (2009).
19. Jones, P. D. et al. Hemispheric and large-scale land-surface air temperature variations: an extensive revision and an update to 2010. *J. Geophys. Res.* **117**, D05127 (2012).
20. Jones, C. D. et al. The HadGEM2-ES implementation of CMIP5 centennial simulations. *Geosci. Model Dev.* **4**, 543–570 (2011).
21. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. A. N. Overview of CMIP5 and the experiment design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
22. Moss, R. H. et al. The next generation of scenarios for climate change research and assessment. *Nature* **463**, 747–756 (2010).
23. Budikova, D. Role of Arctic sea ice in global atmospheric circulation: a review. *Global Planet. Change* **68**, 149–163 (2009).
24. Dethloff, K. et al. A dynamical link between the Arctic and the global climate system. *Geophys. Res. Lett.* **33**, L03703 (2006).
25. Francis, J. A. & Vavrus, S. J. Evidence linking Arctic amplification to extreme weather in mid-latitudes. *Geophys. Res. Lett.* **39**, L06801 (2012).
26. Screen, J. A. & Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* **464**, 1334–1337 (2010).
27. Räisänen, J. CO₂-induced changes in interannual temperature and precipitation variability in 19 CMIP2 experiments. *J. Clim.* **15**, 2395–2411 (2002).
28. Kharin, V. V., Zwiers, F. W., Zhang, X. & Hegerl, G. C. Changes in temperature and precipitation extremes in the IPCC ensemble of global coupled model simulations. *J. Clim.* **20**, 1419–1444 (2007).
29. Katz, R. W. & Brown, B. G. Extreme events in a changing climate: variability is more important than averages. *Clim. Change* **21**, 289–302 (1992).

Supplementary Information is available in the online version of the paper.

Acknowledgements We gratefully acknowledge the Newton Institute for Mathematical Sciences programme Mathematical and Statistical Approaches to Climate Modelling and Prediction. C.H. also acknowledges the CEH National Capability Budget. P.D.J. has been supported by the US DOE (grant DE-SC0005689). V.N.L. and T.M.L. were supported by NERC (grant NE/F005474/1).

Author Contributions C.H. devised the paper and figure structure, performed the numerical analyses and led the interpretation of the results. P.D.J. provided the individual CRU station data. V.N.L. assisted in processing the ECMWF re-analysis products. P.M.C. aided with selection of main results for presentation. T.M.L. and P.M.C. helped to place results in the context of the broader climate change debate. All authors contributed to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.H. (ch@ceh.ac.uk).

METHODS

The ECMWF ERA-40¹¹ gridded data sets are globally complete and have full years of 6-hourly temperature data between 1958 and 2001 inclusive. ECMWF ERA-Interim similarly has 6-hourly data, and at present covers the period from 1979 onwards. As time progressed, the ERA-40 analysis entailed more measurements, most notably satellite retrievals after 1979 (ref. 11). Comparison with the weather-station-based CRUTEM4 data set¹⁹, however, reveals broad agreement, especially concerning year-to-year variability through that transition period^{30,31}. The detrending of data (as we have done, predominantly with a time period of 11 yr) will remove known offsets, which are more present for the Southern Hemisphere and especially the Antarctic¹⁹.

In Fig. 1a, we calculate at each ECMWF grid box the yearly average temperature. Anomalies are then found by initially subtracting from these the average of the yearly temperatures between 1958 and 1980, before then dividing these values by the local standard deviations of these numbers for 1958–1980. For each of the four time periods, we then put all of these normalized yearly anomalies, from across the globe, into bins of size 0.05. To allow area weighting (that is, smaller grid boxes near the poles), bin values are multiplied, for each anomaly value falling in them, by cosine (in degrees) of latitude. The distributions plotted are then scaled to have an area of unity beneath them, thus allowing comparison. Figure 1b is identical to Fig. 1a, except that in Fig. 1b the initial calculation of anomalies for each curve is made relative to the mean of their own time period. We then divide by the local standard deviations of these new anomalies, again for the period 1958–1980.

In Fig. 1c, time-evolving global standard deviations of anomalies are analysed. Here, to remove the longer-term warming signals, anomalies of the ERA-40 data are first created by subtracting from the local, yearly average values their 11-yr running mean. Presented (black curve) for these anomalies and each year of the mid-points of this 11-yr detrending, giving the period 1963–1996, is their area-weighted standard deviation across the globe. We extend these calculations (cyan curve) for the ECMWF ERA-Interim data, which, again with 11-yr detrending, gives a time series for the years 1984–2006 inclusive. Our ERA-40 detrended anomalies are then normalized, at each geographical point, by division by the standard deviation of those numbers for the period 1963–1980. The red curve is the time-evolving, global standard deviation, but now of these normalized anomalies. Figure 1d, e considers grid-box standard deviations before and after 1980, again using the (non-normalized) temperature anomalies relative to 11-yr running means. Standard deviation is calculated as a pair of values for each grid box, and for two periods, 1963–1980 and 1981–1996. In Fig. 1d, we bin, by standard deviation value calculated for the 1963–1980 anomalies, the number of occurrences (area weighted by multiplying by cosine of latitude for each occurrence) of an increase or a decrease in standard deviation in the subsequent period, 1981–1996. Figure 1e is a map of percentage change in standard deviation between these times, again using the same pairs of values calculated at each grid box. Seasonal versions of Fig. 1a, b and e are in Supplementary Information. Furthermore, Supplementary Fig. 3a gives zonal differences in these standard deviations, thus again comparing the

period after 1980 with the period before (plotted at ECMWF latitudinal resolution), and Supplementary Fig. 3b gives zonal temperature changes.

Figure 2 is a plot of time-evolving, latitudinal standard deviation for ERA-40 and ERA-Interim. In each grid box, standard deviations of anomalies are calculated over (moving) 11-yr periods. This, in combination with the initial 11-yr detrending to create the anomalies, gives calculations for the years 1968–1991 inclusive when using ERA-40 and calculations for the years 1989–2001 inclusive when using ERA-Interim. For each of these years, zonal mean standard deviations are then derived for 10° bands and expressed as percentage changes compared to their average value before 1980.

Figure 3a shows four time-evolving, global standard deviations of individual ensemble members of the HadGEM2-ES model, and for historical-plus-RCP8.5 forcings. These are calculated for each year across all model grid boxes (and are area weighted) and again from anomalies that have been derived by local detrending with 11-yr running means. The thick black curve is the mean of the 11-yr moving averages of these four historical-RCP8.5 ensembles. The thick red curve is the 11-yr moving average of the global standard deviation values for ERA-40 and ERA-Interim (and, hence, a smoothed version of the black and cyan curves in Fig. 1b). Figure 3b–e shows zonal averages of 11-yr moving averages of standard deviations from the detrended data, which are hence similar calculations to those of Fig. 2 for the ECMWF data (same colour scheme as Fig. 2, and the change is expressed as a percentage of the mean standard deviation before 1980). For each of the 17 climate models available with simulations made using scenarios of historical-plus-RCP8.5 forcings, and again averaging across ensembles, curves identical to the black one in Fig. 3a are derived and shown in Supplementary Fig. 6. The mean of all of these curves is the black curve in Fig. 3f, which provides an overall time-evolving estimate of global standard deviation averaged across climate models (where, as outlined above, each climate model is averaged across available ensembles of its estimate of 11-yr moving averages of global standard deviation, themselves based on anomalies from 11-yr local detrending). Additionally each model-specific curve in Supplementary Fig. 6 is then scaled such that its mean up to 1980 is the same as the all-model mean. The spread of these curves (time-evolving maximum and minimum global standard deviations across models) is the turquoise area. Figure 3g and Supplementary Fig. 7 are identical to Fig. 3f and Supplementary Fig. 6, respectively, except that local anomalies are instead first detrended with 31-yr running means. We show the moving averages of global standard deviations of these anomalies, with the averaging also now at the 31-yr timescale. All calculations for sea ice (used in Fig. 4 and Supplementary Fig. 8) are based on 31-yr moving averages of annual average areal coverage.

- 30. Simmons, A. J. *et al.* Comparison of trends and low-frequency variability in CRU, ERA-40, and NCEP/NCAR analyses of surface air temperature. *J. Geophys. Res.* **109**, D24115 (2004).
- 31. Simmons, A. J., Willett, K. M., Jones, P. D., Thorne, P. W. & Dee, D. P. Low-frequency variations in surface atmospheric humidity, temperature, and precipitation: Inferences from reanalyses and monthly gridded observational data sets. *J. Geophys. Res.* **115**, D01110 (2010).

The role of behaviour in adaptive morphological evolution of African proboscideans

Adrian M. Lister¹

The fossil record richly illustrates the origin of morphological adaptation through time. However, our understanding of the selective forces responsible in a given case, and the role of behaviour in the process, is hindered by assumptions of synchrony between environmental change, behavioural innovation and morphological response. Here I show, from independent proxy data through a 20-million-year sequence of fossil proboscideans in East Africa, that changes in environment, diet and morphology are often significantly offset chronologically, allowing dissection of the roles of behaviour and different selective drivers. These findings point the way to hypothesis-driven testing of the interplay between habitat change, behaviour and morphological adaptation with the use of independent proxies in the fossil record¹.

Elephants (Elephantidae) arose in Africa from within a paraphyletic assemblage of gomphothere proboscideans (Fig. 1) with low-crowned molars comprising a series of rounded cusps. The fusion of these cusps into transverse 'lophs' led to the lamellar structure of the elephant molar which then, in parallel in different genera, multiplied the number of lamellae and increased the height of the crown (hypodonty)². These changes occurred in Africa against a backdrop of Neogene aridification and replacement of C3-dominated woodland with C4-dominated grassland^{3–5}.

Palaeosol $\delta^{13}\text{C}$ data show that the spread of C4 vegetation in East Africa occurred gradually, beginning around 10 million years (Myr) ago and accelerating after 4 Myr^{4,5} (Fig. 2a). Enamel $\delta^{13}\text{C}$ data (Fig. 2b)

indicate that late gomphotheres and early elephants both switched rapidly from a C3-dominated to a C4-dominated feeding strategy around 8 Myr ago, consuming a substantial proportion of grass at a time when the vegetation was a mosaic still containing significant areas of C3 woodland^{3,4}. This behavioural switch to grazing-dominated mixed feeding occurred in parallel in several proboscidean genera and persisted little-changed through their subsequent history (Fig. 2b). It is interesting, however, that in the derived gomphothere *Anancus* and stegodont *Stegodon*, significant individual or population variation in feeding strategy is seen in their spread of enamel $\delta^{13}\text{C}$ values, exemplifying the behavioural 'choice' that probably originated the switch to grazing at the species level. One group of proboscideans, the deinotheres, remained strict C3-feeders (very probably browsers) within remnant woodland habitat through this period (Supplementary Fig. 1), highlighting the availability of a range of habitats; deinotheres retained a simple, low-crowned dentition throughout⁶.

The interpretation of C3 or C4 feeding in African Neogene ungulates as largely indicating browse or graze, respectively, is supported by cross-taxon studies, including dental mesowear analysis, as well as vegetation data. Equids, in the same East African sites as the proboscideans studied here, show browsing mesowear associated with a C3 enamel $\delta^{13}\text{C}$ signal at 9.9 Myr ago, and grazing mesowear associated with a C4 signal at 7.4 Myr ago³. At Laetoli (Tanzania, ~4.4–2.6 Myr ago), despite phytolith evidence for C3 as well as C4 grasses, the combination of dental $\delta^{13}\text{C}$ and mesowear analysis indicates that the C3

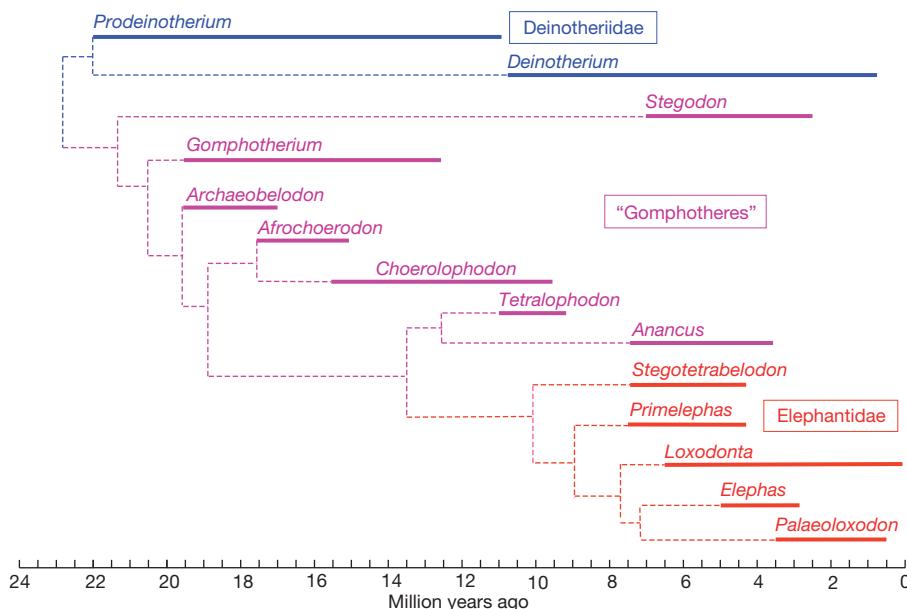


Figure 1 | Phylogeny of proboscidean genera included in this study. Bold lines indicate the approximate duration of existence of each genus in East Africa; *Anancus* and *Palaeoloxodon* endured longer elsewhere in the continent (Supplementary Tables 1–3 and ref. 6). Dashed lines indicate relationships⁶, but

position of nodes does not convey exact date of divergence. *Stegodon* is considered the sister-group of Elephantidae by some authors; others subsume *Palaeoloxodon* within *Elephas* (ref. 6 and Supplementary Information).

¹Department of Earth Sciences, Natural History Museum, London SW7 5BD, UK.

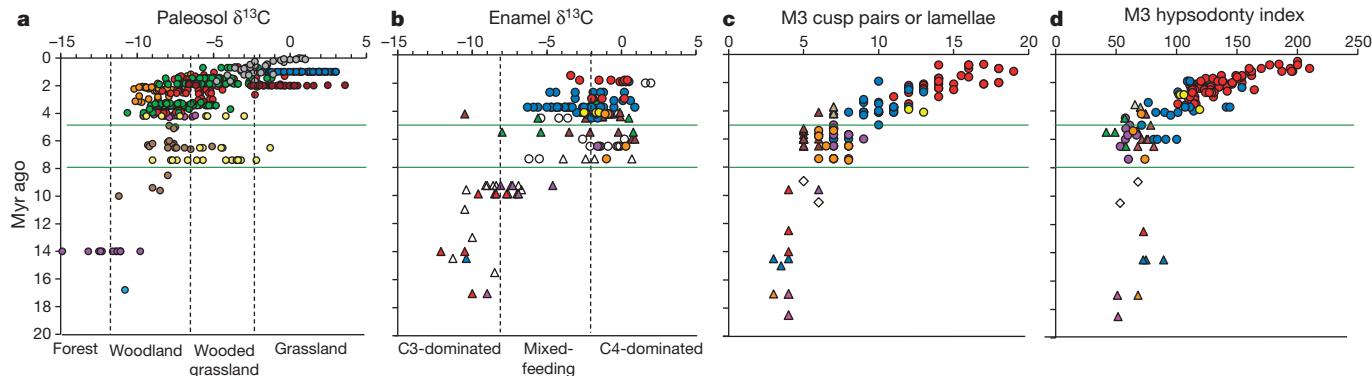


Figure 2 | Environmental, behavioural and morphological trends for East African proboscideans over the past 20 Myr. **a**, The environmental proxy: index of vegetation from palaeosol pedogenic carbonate $\delta^{13}\text{C}$. Data are from selected localities in Kenya, Tanzania (Olduvai), and Ethiopia (Omo Valley). Turquoise, Lothagam; purple, Fort Ternan; brown, Baringo; yellow, Lothagam; pink, Kanapoi; green, Koobi Fora; red, Nachukui; orange, Shungura; maroon, Kanjera; blue, Olorgesailie; grey, Olduvai; see Supplementary Table 1 for details and sources. Vegetational interpretation from ref. 4. **b**, The behavioural (dietary) proxy: $\delta^{13}\text{C}$ from proboscidean dental enamel. Broad dietary classification from ref. 3. **c**, **d**, The adaptive morphology: number of transverse enamel units (cusp pairs, lophs or lamellae) in upper and lower M3

(c), hypsodonty index in upper M3 (**d**). Key to **b–d**: triangles, stegodonts and gomphotheres: pink, *Archaeobelodon*; orange, *Gomphotherium*; blue, *Afrochoerodon*; red, *Choerolophodon*; purple, *Tetralophodon*; brown, *Anancus kenyensis*; pale brown, *A. ultimus*; green, *Stegodon*; pale green, *S. kaisensis* ‘Warwile Stage’; open, indeterminate gomphothere. Circles, elephants: orange, *Stegotetrabelodon*; purple, *Primelephas*; blue, *Loxodonta*; yellow, *Elephas*; red, *Palaeoloxodon*; open, indeterminate elephant. Open diamonds: tetralophodont/elephantid intermediate. Horizontal green lines mark interval 8–5 Myr ago when diet changed (**b**) but with little morphological response (**c**, **d**). See Supplementary Tables 2 and 3 for details of species, localities and dating. Supplementary Fig. 2 provides an enlarged version of this figure.

component of a range of mixed-feeding mammal species was largely browse⁷. Nonetheless, it cannot be ruled out that some C3-feeding proboscideans before 8 Myr ago may have been taking a proportion of C3 grass (like modern elephants that favour bamboo, a C3 grass⁸).

The Late Miocene transition from C3 browsing to C4 grazing is seen in other ungulate groups in Africa and elsewhere^{3,9}, and has been linked to an increase in the crown height (hypodonty) of the molars. Elephants underwent a threefold increase in crown height (Fig. 2d) as well as substantial multiplication of the number of enamel lamellae (Fig. 2c), both considered adaptations to increased dietary wear, and in the case of lamellar number, possibly also to the mechanics of chewing grass. Strikingly, most of the increase in hypodonty and lamellar number occurred after 5 Myr ago, a lag of at least 3 Myr after the adoption of a grass-dominated diet (Fig. 2). Small increases in lamellar number are seen at around 10 Myr ago and, together with modest hypodonty increase, at 8–6 Myr ago, coincident with shifts in $\delta^{13}\text{C}$ both in pedogenic carbonate and tooth enamel (dietary signal); we lack the resolution to assess whether these anatomical shifts trailed the behavioural shift. However, the limited nature of this dental response to grazing among the bunodont gomphotheres and brachydont early elephants and stegodonts before 5 Myr ago (compare Fig. 2b and 2d) is surprising and remarkable, as these are morphologies generally associated with soft-leaf browsing¹⁰.

Such offsets between behaviour and morphology demand that we think beyond the traditional palaeontological assumption of a lock-step between the two. One possibility is that behavioural accommodation preceded morphological adaptation by placing the organism in a new niche where selection then led to morphological adaptation¹. The behavioural switch to C4 grazing at around 8 Myr ago in late gomphotheres and early elephants would have led to selection for enhanced dental resistance to abrasion. Grass contains more biogenic silica (phytoliths) than browse¹¹, and grazing (generally low-feeding) ungulates also pick up more soil-derived grit than tree- and shrub-feeders^{12–14}. Even if some proboscideans were taking a proportion of C3 grass before this date, this might not have produced strong selection on dental morphology because of the lower phytolith concentration than in C4 grasses¹⁵ and the semi-wooded habitat with limited pedogenic and airborne dust (see below). Early responses to grazing pressure are seen in the minor dental advances before 5 Myr ago. But the major adaptation, a substantial increase in hypodonty, required significant reorganization of cranial

anatomy to accommodate high-crowned teeth², which could account, in part, for the marked delay in its appearance.

The concept of an evolutionary ‘lag’ due to behavioural initiation has sometimes been considered implausible because of the implied period of ‘maladaptation’¹³. If the gomphotheres, stegodonts and elephants in the 8–5 Myr ago interval were suffering enhanced tooth wear as a result of grazing, their survival could have been ensured by countervailing competitive advantages of exploiting a new niche, but with life-history adjustment to reduced longevity (for example, earlier reproduction) as their teeth wore out sooner. Increased tooth-life through hypodonty and lamellar increase, reversing this trend, would then have been at a selective premium. Studies of several lineages of North American mammals also indicate significant time lag between the appearance of open habitats and evolution of hypodonty^{1,15}. For example, in a study of horses (Equidae) with dental mesowear as dietary proxy, a gradual transition to grass-eating began around 22 Myr ago, but major hypodonty increase did not begin until about 14 Myr ago; populations in the intervening period are considered to have been pioneering new habitats and consequently under elevated selection pressure^{14,16}. Another morphological feature in proboscideans potentially explaining their adaptation to graze after 8 Myr ago, is their peculiar system of horizontal dental replacement. However, this had already largely been achieved before that time, providing a large lifetime dental battery for feeding very large body size, within a short jaw biomechanically linked to the heavy head and tusks². More significantly, the advent around 7 Myr ago of a block-like molar crown of parallel lamellae held together by cement was the key dental adaptation of crown group Elephantidae that subsequently allowed increased hypodonty and lamellar multiplication, explaining their ultimate success at the expense of the gomphotheres and stegodonts.

A second factor potentially explaining the ‘lag’ between the onset of grass-eating and the morphological response is a decoupling of the selective factors associated with grazing. Whereas the switch to grass-eating is attested by the strong C4 signal within enamel $\delta^{13}\text{C}$ after 10–8 Myr ago, the quantity of soil-derived grit might have increased later. We do not yet have good proxy data for environmental grit, or its consumption by fossil mammals, although dental microwear is a potential tool¹⁷. However, the remarkable parallel, from 3.5 Myr ago onwards, between the increasingly open environments indicated by palaeobotanical evidence and palaeosol $\delta^{13}\text{C}^{4,18}$ (Fig. 2a), and the

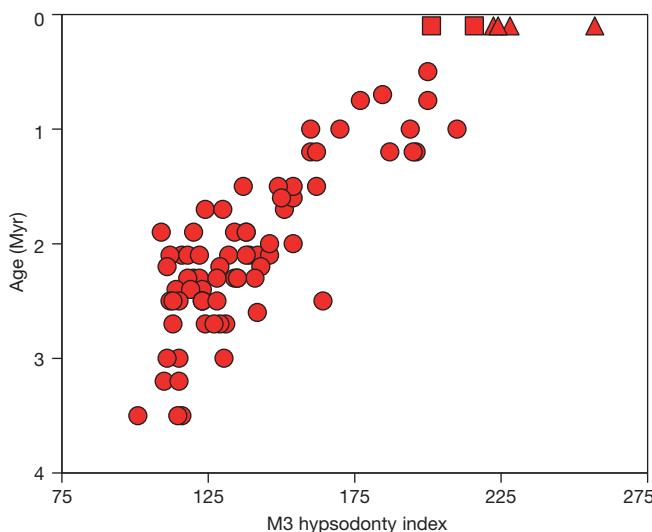


Figure 3 | Incremental evolution of hypsodonty index in the African Plio-Pleistocene *Palaeoloxodon recki/iolensis* lineage. Hypsodonty index from data in Fig. 2d (circles) plus *P. iolensis* from North Africa (squares) and southern Africa (diamonds).

major trend in elephant hypsodonty and lamellar number (Fig. 2c, d), could be due to additional selective pressure imposed by increasing quantities of dust and grit, both airborne and from soil at the place of feeding. This in turn was driven by the increased aridity after 3.5 Myr ago¹⁹, a factor known to be strongly correlated with airborne particulate matter²⁰, combined with the absence of local tree cover acting as sediment traps²¹. The pressure would have been enhanced by the shorter grass species typical of more arid grassland, leading to increased soil uptake by grazing ungulates²², and by competition from the expanding guild of grazing species²³.

These factors imply an evolutionary ‘ratchet effect’ whereby the initial behavioural commitment to grass-eating at 10–8 Myr ago not only placed species in a niche with selection for morphological adaptation, but where later environmental changes increased the selective pressure. The resulting trend in hypsodonty and lamellar number in the *Palaeoloxodon recki/iolensis* lineage (= *Elephas recki/iolensis* of some authors), in particular, represents a remarkable incremental transformation^{2,23} (Fig. 3). Maximal values of lamellar number and hypsodonty are not reached until 0.5 Myr ago, at least 7 Myr after the onset of a grazing-dominated diet. Parallel shifts in morphology, with very similar chronology, are seen in two other elephant lineages originating in Africa, *Elephas*²⁴ and *Mammuthus*²⁵, in the different environmental settings of southern and northern Eurasia, respectively. This might reflect similar environmental pressures, or a common limit to the rate of adaptation following an initial shift to grazing. A further hypothesis for the protracted evolution of hypsodonty, an evolutionary arms race between plant phytoliths and mammalian dental adaptation, is not supported by phytolith data through the history of C4 grasses²⁶.

In contrast to the evolutionary trends in elephantids, the two late elephantoid lineages *Stegodon kaisensis* and *Anancuskenensis/ultimus*, having survived a behavioural shift to grazing, lacked the key dental innovations allowing hypsodonty, so apparently succumbed to further environmental change. Both show the addition of an enamel cusp pair or loph (Fig. 2c) and complexification of the enamel band around 4 Myr ago^{6,27,28}, a ‘last throw’ to achieve increased resistance to abrasion in molars, but evidently insufficient as they went regionally extinct soon after⁶.

Finally, this study sheds light on the origin of the living African forest and savannah elephants *Loxodonta cyclotis* and *Loxodonta africana* (*L. a. africana* and *L. a. cyclotis* of some authors). Their similar hypsodonty index (typically in the range 1.7–2.1) and lamellar count (mode 10–12) (see Methods) indicates a semi-open, mixed-feeding

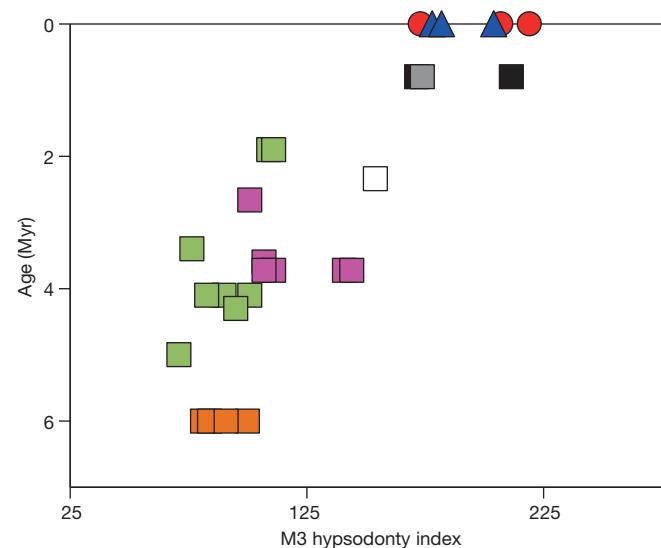


Figure 4 | Hypsodonty evolution in African Pliocene to Recent *Loxodonta*. Hypsodonty data from Fig. 2d plus southern and North African *L. atlantica* and Recent *L. africana* and *L. cyclotis*. Squares, fossil samples: orange, *L. cookei*; green, *L. adaaurora*; purple, *L. exoptata*; white, *L. sp.*; grey, *L. atlantica zulu*; black, *L. atlantica atlantica*. Blue triangle, *L. africana*; red circle, *L. cyclotis*.

ancestry for *L. cyclotis* despite its current forest habitat and browsing diet. Fossil evidence shows that whereas modern lamellar count was attained in the lineage by about 3.5 Myr ago, modern hypsodonty index was not reached until after 2 Myr ago (Figs 2c, d and 4), so *Loxodonta* may have entered Central African forests around this time, consonant with its dietary flexibility shown by a wide range of $\delta^{13}\text{C}$ values (Fig. 2b), and the lack of its remains on the savannahs after this date, as *Palaeoloxodon recki* became ecologically dominant²³. These data also indicate that the divergence date between *L. africana* and *L. cyclotis* was at, or somewhat more recent than, the younger end of the range 2.6–5.6 Myr ago estimated from nuclear DNA²⁹. Whether the population ancestral to *L. africana* remained in the savannah throughout, or re-emerged from the forest³⁰, remains to be seen.

METHODS SUMMARY

Palaeosol carbonate and dental enamel $\delta^{13}\text{C}$ values were collated from the literature and are expressed as ‰ relative to standard. Because of similar fractionation relative to the living plant (~14‰ for enamel⁹, 12–17‰ for pedogenic carbonate⁴), they occupy similar but not directly comparable ranges. Interpretation of $\delta^{13}\text{C}$ ranges as vegetation types is based on modern East African data from ref. 4, and as dietary categories from ref. 3. Only East African proboscidean genera for which $\delta^{13}\text{C}$ enamel data are available are included in the morphometric analysis. Taxa are plotted as genera; species designations and taxonomic revisions are given in Methods and Supplementary Tables 2 and 3. Dental parameters are calculated from literature data and original measurements. Hypsodonty index is 100 × maximum unworn height of upper M3 crown divided by maximum crown width including cement. Transverse cusp pairs in gomphotheres, lophs in *Stegodon* and early elephants, and lamellae (plates) in later elephants are considered homologous; their number excludes ‘talons’ and is pooled for upper and lower M3. Sample ages are based mainly on radiometrically dated volcanic tuffs (Supplementary Tables 1–3).

Full Methods and any associated references are available in the online version of the paper.

Received 21 February; accepted 8 May 2013.

Published online 26 June; corrected online 14 August 2013 (see full-text HTML version for details).

1. Lister, A. M. Behavioural leads in evolution – evidence from the fossil record. *Biol. J. Linn. Soc.* (in press).
2. Maglio, V. J. Origin and evolution of the Elephantidae. *Trans. Am. Phil. Soc.* **63**, 1–149 (1973).

3. Uno, K. T. *et al.* Late Miocene to Pliocene carbon isotope record of differential diet change among East African herbivores. *Proc. Natl Acad. Sci. USA* **108**, 6509–6514 (2011).
4. Cerling, T. E. *et al.* Woody cover and hominin environments in the past 6 million years. *Nature* **476**, 51–56 (2011).
5. Levin, N. E., Brown, F. H., Behrensmeyer, A. K., Bobe, R. & Cerling, T. E. Paleosol carbonates from the Omo Group: isotopic records of local and regional environmental change in East Africa. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **307**, 75–89 (2011).
6. Sanders, W. J., Gheerbrant, E., Harris, J. M., Saegusa, H. & Delmer, C. in *Cenozoic Mammals of Africa*. (eds Werdelin, L. & Sanders, W. J.) 161–251 (Univ. of California Press, 2010).
7. Kingston, J. D. in *Paleontology and Geology of Laetoli: Human Evolution in Context Volume 1: Geology, Geochronology, Paleoecology and Paleoenvironment* (ed. T. Harrison) Ch. 15 293–328 (Springer, 2011).
8. Cerling, T. E., Harris, J. M. & Leakey, M. G. Browsing and grazing in elephants: the isotope record of modern and fossil proboscideans. *Oecologia* **120**, 364–374 (1999).
9. Cerling, T. E. *et al.* Global change through the Miocene/Pliocene boundary. *Nature* **389**, 153–158 (1997).
10. Zazzo, A. *et al.* Herbivore paleodiet and paleoenvironmental changes in Chad during the Pliocene using stable isotope ratios of tooth enamel carbonate. *Paleobiology* **26**, 294–309 (2000).
11. Hummel, J. *et al.* Another one bites the dust: faecal silica levels in large herbivores correlate with high-crowned teeth. *Proc. R. Soc. Lond. B* **278**, 1742–1747 (2011).
12. Mendoza, M. & Palmqvist, P. Hypsodonty increase in ungulates: an adaptation for grass consumption or for foraging in open habitat? *J. Zool.* **274**, 134–142 (2008).
13. Janis, C. M. in *The Ecology of Browsing and Grazing* (eds Gordon, I. J. & Prins, H. H. T.) 21–45 (Springer, 2008).
14. Damuth, J. & Janis, C. M. On the relationship between hypsodonty and feeding ecology in ungulate mammals, and its utility in palaeoecology. *Biol. Rev. Camb. Philos. Soc.* **86**, 733–758 (2011).
15. Strömberg, C. A. E. Evolution of hypsodonty in equids: testing a hypothesis of adaptation. *Paleobiology* **32**, 236–258 (2006).
16. Mihlbachler, M. C., Rivals, F., Solounias, N. & Semprebon, G. M. Dietary change and evolution of horses in North America. *Science* **331**, 1178–1181 (2011).
17. Lucas, P. W. *et al.* Mechanisms and causes of wear in tooth enamel: implications for hominin diets. *J. R. Soc. Interface* **10**, 20120923 (2013).
18. Strömberg, C. A. E. Evolution of grasses and grassland ecosystems. *Annu. Rev. Earth Planet. Sci.* **39**, 517–544 (2011).
19. Feakins, S. J. & DeMenocal, P. B. in *Cenozoic Mammals of Africa* (eds Werdelin, L. & Sanders, W. J.) 45–55 (Univ. of California Press, 2010).
20. Valentin, C. in *Global Change and Terrestrial Ecosystems* (eds Walker, B. H. & Steffen, W.) 317–338 (Cambridge Univ. Press, 1996).
21. Bealey, W. J. *et al.* Estimating the reduction of urban PM₁₀ concentrations by trees within an environmental information system for planners. *J. Envtl. Mgmt.* **85**, 44–58 (2007).
22. Codron, D. *et al.* Functional differentiation of African grazing ruminants: an example of specialized adaptations to very small changes in diet. *Biol. J. Linn. Soc.* **94**, 755–764 (2008).
23. Sanders, W. J. & Haile-Selassie, Y. A new assemblage of Mid-Pliocene proboscideans from the Woranso-Mille Area, Afar Region, Ethiopia: taxonomic, evolutionary, and paleoecological considerations. *J. Mamm. Evol.* **19**, 105–128 (2012).
24. Lister, A. M. *et al.* New fossil remains of *Elephas* from the southern Levant: implications for the evolutionary history of the Asian elephant. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* <http://dx.doi.org/10.1016/j.palaeo.2013.05.013> (20 May 2013).
25. Lister, A. M. & Sher, A. V. Gradual evolution and speciation in the origin of the woolly mammoth. *Science* **294**, 1094–1097 (2001).
26. Strömberg, C. A. E. The origin and spread of grass-dominated ecosystems in the Late Tertiary of North America: preliminary results concerning the evolution of hypsodonty. *Palaeogeogr. Palaeoclimatol. Palaeoecol.* **177**, 59–75 (2002).
27. Hautier, L. *et al.* New material of *Anancus kenyensis* (Proboscidea, Mammalia) from Toros-Menalla (Late Miocene, Chad): contribution to the systematics of African anancines. *J. Afr. Earth Sci.* **53**, 171–176 (2009).
28. Sanders, W. J. in *Paleontology and Geology of Laetoli: Human Evolution in Context Volume 2: Fossil Hominins and the Associated Fauna* (ed. Harrison, T.) 233–262 (Springer, 2011).
29. Rohland, N. *et al.* Proboscidean mitogenomics: chronology and mode of elephant evolution using mastodon as outgroup. *PLoS Biol.* **5**, e207 (2007).
30. Eggert, L. S., Rasner, C. A. & Woodruff, D. S. The evolution and phylogeography of the African elephant inferred from mitochondrial DNA sequence and nuclear microsatellite markers. *Proc. R. Soc. Lond. B* **269**, 1993–2006 (2002).

Supplementary Information is available in the online version of the paper.

Acknowledgements I thank T. E. Cerling, W. J. Sanders and C. M. Janis for discussion; K. Uno, Y. Kunimatsu and H. Nakaya for assistance with specimen identification; and P. Hadland, R. Portela-Miguez and W. Wendelin for access to collections.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The author declares no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.L. (a.lister@nhrm.ac.uk).

METHODS

Palaeosol and enamel stable isotopes. Palaeosol carbonate and dental enamel $\delta^{13}\text{C}$ values were collated from the literature (Supplementary Tables 1 and 2 and references in Supplementary Information), and are expressed as ‰ relative to PDB (Pee Dee Belemnite) standard. Because of similar fractionation relative to the living plant (~14‰ for enamel³, 12–17‰ for pedogenic carbonate⁴), they occupy similar but not directly comparable ranges. Interpretation of $\delta^{13}\text{C}$ ranges as vegetation types is based on modern East African data from ref. 4, and as dietary categories from ref. 3.

Palaeosol carbonate $\delta^{13}\text{C}$ in Fig. 2a are from the Turkana basin and other localities in Kenya and Tanzania (Supplementary Table 1). Large numbers of determinations are available in the published sources³¹; where appropriate, multiple determinations from the same sample or in some cases same horizon have been averaged. Fig. 2a shows considerable inter-site variation in the $\delta^{13}\text{C}$ signal, and hence implied vegetation, even between regionally close localities (exemplified by those around Lake Turkana). However, it is notable that despite the offset in absolute values between localities, the trend to reduced $\delta^{13}\text{C}$ after 4 Myr ago is seen in each Formation (Koobi Fora, East Turkana; Nachukui, West Turkana; Shungura, Omo Valley). Additional determinations have been published for some of these localities and corroborate the pattern in Fig. 2a. Examples include Bed I at Olduvai (~1.8 Myr ago), with $\delta^{13}\text{C}$ values in the region –4 to –6‰ (ref. 32), and additional data from Koobi Fora in the interval 2.0–1.5 Myr ago corroborating a $\delta^{13}\text{C}$ trend from about –7 to about –5‰ (ref. 33). Further afield, a large set of palaeosol carbonate $\delta^{13}\text{C}$ values from the Awash valley, Ethiopia shows some differences from the Omo-Turkana Basin, but with the same broad trend of opening habitat over the past 4 Myr³⁴, indeed with more positive $\delta^{13}\text{C}$ values than Turkana at some points in the sequence. Deposits in the region of 19–18 Myr ago at Rusiga Island (Kenya) have pedogenic carbonate $\delta^{13}\text{C}$ values clustered around –8 to –10‰, indicating woodland as expected for this time interval but more positive than at Fort Ternan (~14 Myr ago), interpreted as reflecting a C3 vegetation with reduced photosynthetic fractionation due to a semi-arid episode evidenced by sedimentology³⁵. Another series of early samples with relatively positive $\delta^{13}\text{C}$ values (about –5.2 to –7.4‰) came from the Muruyur Formation of the Tugen Hills, and was interpreted as reflecting a mixed C3/C4 vegetation as early as 15 Myr ago³⁶. However, these results have been disputed in terms of the origin and age of the analysed carbonates and other palaeobotanical and faunal evidence indicating forest in the region at the time^{37,38}. Somewhat later, the wide range of values at Lothagam, South Turkana (Fig. 2a) implies a local expansion of relatively open (wooded grassland) habitat, earlier than in other areas³⁹.

Despite the general trend towards increasingly open environments through the Pliocene, palaeobotanical and faunal evidence indicates that African environments retained a mosaic character throughout the late Cenozoic, with some locally wooded or even forested areas even after the spread of grassland⁴⁰. The Proboscidean family Deinotheriidae retained a strongly browsing $\delta^{13}\text{C}$ enamel signature throughout the succession (Supplementary Fig. 1), implying persistence within a woodland habitat. This is seen both in *Prodeinotherium* in the interval 18–12 Myr ago, with $\delta^{13}\text{C}$ values similar to those of contemporary gomphotheres, and its likely descendent *Deinotherium* in the interval 5–1 Myr ago, when contemporary gomphotheres and elephants, often preserved at the same localities (Supplementary Table 2), show strongly mixed-feeding to grazing diets.

Morphometrics, taxonomy and identification. Only East African proboscidean species for which $\delta^{13}\text{C}$ enamel data are available are included in the morphometric analysis. Dental parameters are calculated from literature data and original measurements. Hypsodonty index is $100 \times$ maximum unworn height of upper M3 crown divided by maximum crown width including cement². Transverse cusp pairs in gomphotheres, lophs in *Stegodon* and early elephants, and lamellae (plates) in later elephants are considered homologous; their number excludes ‘talons’ and is pooled for upper and lower M3. Sample ages are based mainly on radiometrically-dated volcanic tuffs (Supplementary Table 3).

Proboscidean taxa are plotted by genus in Fig. 2, but individual species (and in some cases subspecies) designations can be found in Supplementary Tables 2 and 3. For most genera, only a single species occurs in East Africa in the time period considered. Exceptions are included in the following alphabetical list of taxonomic notes and revisions:

Afrochoerodon. The trilophodont gomphothere *A. kisumuensis* was described by MacInnes⁴¹ as *Trilophodon angustidens kisumuensis*. It was recognized as a choerolophodont by Tassy⁴² who named it *Choerolophodon kisumuensis*, and placed in *Afrochoerodon* by Pickford⁴³.

Anancus. Most anancines are allocated to *A. kenyensis* (MacInnes, 1942) except for some advanced specimens recently allocated to *Anancus ultimus* Sanders, 2011 (ref. 28). These are marked separately in Fig. 2c.

Anancus/Choerolophodon. A specimen from Nakali, Kenya, dated to 10.0–8.8 Myr, has been listed and discussed as *Anancus*^{7,8}. However, this is considerably earlier than the otherwise known entry of *Anancus* into Africa^{6,44}, and in a revised faunal list for the site only *Deinotherium*, *Choerolophodon* and indeterminate Elephantidae have been identified⁴⁵. Y. Kunimatsu (personal communication, 2011) indicates that the specimen is a small fragment, not fully identifiable but by analogy with other Nakali fauna probably *Choerolophodon*.

Gomphotheriidae indet. Specimens plotted as ‘Gomphotheriidae indet.’ in Fig. 2b are fragments of enamel, mainly from Lothagam (Lower Nawata) and Namurungule. For the Lower Nawata, *Anancus kenyensis* (tetra/pentalophodont) is the only gomphothere identified⁴⁶, whereas for Namurungule, only *Tetralophodon* and *Choerolophodon* have been reported⁴⁷. These are therefore likely identifications but other taxa cannot be excluded.

Mammuthus. Mastodons are not included in the present study since, although they show striking evolution of hypsodonty and lamellar number through the Plio-Pleistocene, most of this evolution took place in Eurasia⁴⁸. There is also at present no enamel isotopic data on African *Mammuthus*. Our current limited knowledge of African remains referred to this genus is summarized in ref. 23.

Palaeoloxodon is represented by *P. recki*, often named *Elephas recki* Dietrich, 1915, but here considered congeneric with Eurasian species of *Palaeoloxodon* Matsumoto^{49,50}; others prefer subspecific status as *E. (P.) recki*⁵¹. This taxonomic issue will not be resolved without a cladistic analysis of the whole *Elephas-Palaeoloxodon* clade. A series of subspecies of *P. recki* has been erected^{52,53}, and specimens have been listed by their allocated subspecies in Supplementary Tables 2 and 3. Some authors^{54,55} have suggested abandoning these names because their morphological and temporal ranges overlap, while others⁵⁶ retain them as useful labels for biochronological stages. There is also disagreement on whether the oldest member, *E. r. brumpti*, should be included in the lineage^{51,56}. Data have been published² on some *P. recki* molars additional to those tabulated here^{52,57}, but these have not been included as their localities and horizons (and hence ages) were not individually noted². Some NHM specimens listed in ref. 2 are from early 20th century excavations at Kanjera, and hence of uncertain age⁵⁸, or localized only to ‘Homa Mountain’. The species *P. iolensis* (Pomel, 1895), known mainly from North and southern Africa, is considered the terminal member of the African *Palaeoloxodon* lineage^{2,6}. The term ‘incremental’ used to describe the morphological trajectory is intended to convey change through a series of stages; it is preferred to ‘gradual’ which suggests perfect continuity or infinitesimal gradations. Nor does the observed incremental change necessarily imply a purely anagenetic mode of evolution: it is possible that successive stages evolved in isolated populations, that is, as cladogenetic events at or below the species level.

Primelephas, *Stegotetrabelodon* and Elephantidae indet. Specimens plotted as ‘Elephantidae indet.’ in Fig. 2b are fragments of enamel, mainly from Lothagam and the Adu-Asa Formation (Awash), likely to be either *Primelephas* or *Stegotetrabelodon*. *P. korotorensis* Coppens, 1965 includes *P. gomphotheroides* Maglio, 1970 since they are likely synonyms, *P. korotorensis* taking priority⁶. Finally, molar LT23783 from the Upper Nawata of Lothagam is the holotype of *Elephas nawataensis* Tassy 2003 (ref. 46), but has been referred by other authors to *Primelephas*⁶.

Stegodon. This genus, known mainly from Eurasia, is of uncertain systematic position; there is debate whether its elephant-like features are synapomorphic or convergent^{6,59}. If convergent, *Stegodon* is likely to be basal to Gomphotheriidae (as shown in Fig. 1), and Elephantidae derived instead from advanced tetralophodont gomphotheres such as the specimens from Kakara and Chorora (see below).

Tetralophodon, and Tetralophodont/Elephantid intermediates. Specimens plotted under *Tetralophodon* are tetrabelodont gomphotheres *sensu lato*. There is debate over the allocation of some molars in the 10–9 Myr ago interval that are agreed to be morphologically at the borderline between tetralophodonts and elephantids. These include a molar from the Kakara Formation of Uganda, described as a ‘perfect intermediate between a gomphothere and an elephant’⁶⁰, and a molar from Chorora (Awash), considered an advanced tetralophodont gomphothere close to the ancestry of elephants^{61,62}. As their attribution to gomphotheres or elephantids is debated, they have been plotted in Fig. 2 as ‘tetralophodont/elephantid intermediates’.

Additional minor notes on taxonomy and identification are given in Supplementary Tables 2 and 3.

Dental evolution in Loxodonta. The living forest and savannah forms of African elephant are currently formally recognized as subspecies, *L. africana cyclotis* and *L. africana africana*, respectively⁶³, but their distinctiveness based on nuclear DNA has led to suggestions that they should be recognized as separate species⁶⁴. Here they are named as species for clarity, although dentally, apart from the smaller size of *L. cyclotis*, they are indistinguishable. The number of lamellae (excluding talons) in M3 ranges from 9–13 (mode 10–12) in both forms (refs 65–67 and Supplementary Table 3). Hypsodonty index in *L. africana* is typically in the range 1.7–2.1 (refs 2,68 and Supplementary Table 3). There is little comparative data for

L. cyclotis, but two molars from Congo (RMCA 18498 and un-numbered, Royal Museum for Central Africa, Belgium) have hypsodonty indices of 1.73 and 2.07, respectively (data courtesy of W. Wendelen), and a third (Powell-Cotton Museum 42) has a hypsodonty index of ~2.19. All of these fall within the *L. africana* range.

Suggestions (without data) that *L. cyclotis* is lower-crowned than *L. africana*^{6,9} are probably due either to observation of molars partly buried in skulls, or to consideration of crown height without reference to other dimensions of the tooth.

In the fossil record, two distinct loxodontine lineages are recognized: on the one hand *L. adaurora* Maglio, 1970, and on the other *L. cookei* Sanders, 2007, *L. exoptata* (Dietrich, 1941) and *L. atlantica* (Pomel, 1879), the latter group considered more closely related to the living species^{6,28}. In Fig. 2d the earliest points, at 6.0 Myr ago, are *L. cookei*; and of the later points, *L. exoptata* clusters with higher hypsodonty index values than *L. adaurora* (see Fig. 4).

Lamellar counts similar to modern appear in *L. exoptata* from the Upper Laetoli Beds at ~3.85–3.60 Myr ago^{28,70} (Supplementary Table 3, Fig. 2c), but M3 hypsodonty index in the range 100–150, lower than modern, persists until at least 2 Myr ago when the East African record ends (Fig. 2d). Therefore, unless we are missing earlier, more advanced populations, the completion of hypsodonty increase to modern *L. africana* and *L. cyclotis* values occurred after 2 Myr ago. The deduction that the progenitors of *L. cyclotis* must have entered their present niche after this date is based on the assumption that the degree of dental advancement seen in *L. cyclotis* (equivalent to *L. africana*) would not have evolved in a browser living in canopy forest. However, the paucity of fossils from the equatorial forest zone makes direct testing of this hypothesis difficult.

There are no later fossils directly relevant to the origin and divergence of the two living forms. *L. atlantica* from the Middle to Late Pleistocene of North and southern Africa shows large body size and somewhat higher lamellar count (13–14) than the living species, but similar hypsodonty index (1.7–2.1, Fig. 4, Supplementary Table 3). However, details of its enamel loop structure appear to preclude ancestry to living savannah elephants^{2,6}.

31. Levin, N. E. Compilation of East Africa soil carbonate stable isotope data. *Integrated Earth Data Applications* <http://dx.doi.org/10.1594/IEDA/100231> (2013).
32. Sikes, N. E. & Ashley, G. M. Stable isotopes of pedogenic carbonates as indicators of paleoecology in the Plio-Pleistocene (upper Bed I), western margin of the Olduvai Basin, Tanzania. *J. Hum. Evol.* **53**, 574–594 (2007).
33. Quinn, R. L., Lepre, C. J., Wright, J. D. & Feibel, C. S. Paleogeographic variations of pedogenic carbonate $\delta^{13}\text{C}$ values from Koobi Fora, Kenya: implications for floral compositions of Plio-Pleistocene hominin environments. *J. Hum. Evol.* **53**, 560–573 (2007).
34. Levin, N. E., Quade, J., Simpson, S. W., Semaw, S. & Rogers, M. Isotopic evidence for Plio-Pleistocene environmental change at Gona, Ethiopia. *Earth Planet. Sci. Lett.* **219**, 93–110 (2004).
35. Bestland, E. A. & Krull, E. S. Palaeoenvironments of Early Miocene Kisingiri volcano *Proconsul* sites: evidence from carbon isotopes, palaeosols and hydromagmatic deposits. *J. Geol. Soc. Lond.* **156**, 965–976 (1999).
36. Kingston, J. D., Marino, B. D. & Hill, A. Isotopic evidence for Neogene hominid paleoenvironments in the Kenya Rift Valley. *Science* **264**, 955–959 (1994).
37. Cerling, T. E. *et al.* Expansion and emergence of C4 plants. *Nature* **371**, 112–113 (1994).
38. Ségalen, L., Lee-Thorp, J. A. & Cerling, T. Timing of C4 grass expansion across sub-Saharan Africa. *J. Hum. Evol.* **53**, 549–559 (2007).
39. Cerling, T., Harris, J. & Leakey, M. in *Lothagam: The Dawn of Humanity in Eastern Africa Vol. 1* (eds Leakey, M. G. & Harris, J. M.) 605–624 (Columbia Univ. Press, 2003).
40. Bobe, R. The evolution of arid ecosystems in eastern Africa. *J. Arid Environ.* **66**, 564–584 (2006).
41. MacInnes, D. G. Miocene and post-Miocene Proboscidea from East Africa. *Trans. Zool. Soc. Lond.* **25**, 33–106 (1942).
42. Tassy, P. Présence du genre *Choerolophodon* Schlesinger (Proboscidea, Mammalia) dans le Miocène est-africain. *C. r. hebd. Séanc. Acad. Sci. Paris* **284**, 2487–2490 (1977).
43. Pickford, M. *Afrochoerodon* nov. gen. *kisumuensis* (MacInnes) (Proboscidea, Mammalia) from Cheparawa, Middle Miocene, Kenya. *Ann. Paleontol.* **87**, 99–117 (2001).
44. Tassy, P. Nouveaux *Elephantoidea* (Mammalia) dans le Miocène du Kenya. *Cahiers de Paléontologie* (CNRS, Paris, 1986).
45. Kunimatsu, Y. *et al.* A new Late Miocene great ape from Kenya and its implications for the origins of African great apes and humans. *Proc. Natl. Acad. Sci. USA* **104**, 19220–19225 (2007).
46. Tassy, P. in *Lothagam: The Dawn of Humanity in Eastern Africa* (eds Leakey, M. G. & Harris, J. M.) 331–358 (Columbia Univ. Press, 2003).
47. Tsujikawa, H. The palaeontology of *Samburupithecus kiptalami* based on its associated fauna. *Afr. Study Monogr.* **32** (Suppl.), 51–62 (2005).
48. Lister, A. M., Sher, A. V., van Essen, H. & Wei, G. The pattern and process of mammoth evolution in Eurasia. *Quat. Int.* **126–128**, 49–64 (2005).
49. Inuzuka, N. & Takahashi, K. Discrimination between the genera *Palaeoloxodon* and *Elephas* and the independent taxonomical position of *Palaeoloxodon* (Mammalia: Proboscidea). *Misclánea en homenaje a Emiliano Aguirre* **2**, 234–244 (2004).
50. Shoshani, J. *et al.* Relationships within the Elephantinae using hyoid characters. *Quat. Int.* **169–170**, 174–185 (2007).
51. Saegusa, H. & Gilbert, W. H. in *Homo erectus in Africa, Pleistocene Evidence from the Middle Awash* (eds Henry, W., Gilbert, W. H. & Asfaw, B.) 193–226 (Univ. of California Press, 2008).
52. Beden, M. *Les Elephants (Loxodonta et Elephas) d'Afrique Orientale. Systématique, phylogénie, intérêt biochronologique*. PhD thesis, Univ. Poitiers (1979).
53. Beden, M. *Elephas recki* Dietrich, 1915 (Proboscidea, Elephantidae). Évolution au cours du Plio-Pléistocène en Afrique orientale. *Geobios* **13**, 891–901 (1980).
54. Todd, N. E. Reanalysis of African *Elephas recki*: implications for time, space and taxonomy. *Quat. Int.* **126–128**, 65–72 (2005).
55. Ferretti, M. P., Ficcarelli, G., Libsekal, Y., Tecle, T. M. & Rook, L. Fossil elephants from Buia (Northern Afar Depression, Eritrea) with remarks on the systematics of *Elephas recki* (Proboscidea, Elephantidae). *J. Vertebr. Paleontol.* **23**, 244–257 (2003).
56. Sanders, W. J. & Haile-Selassie, Y. A new assemblage of Mid-Pliocene proboscideans from the Worano-Mille Area, Afar Region, Ethiopia: taxonomic, evolutionary, and paleoecological considerations. *J. Mamm. Evol.* **19**, 105–128 (2012).
57. Beden, M. in *Koobi Fora Research Project Vol. 2* (ed. Harris, J. M.) 40–129 (Clarendon Press, 1983).
58. Behrensmeyer, A. K. *et al.* The Pleistocene locality of Kanjera, Western Kenya: stratigraphy, chronology and paleoenvironments. *J. Hum. Evol.* **29**, 247–274 (1995).
59. Kalb, J. E. & Mebrate, A. Fossil elephantoids from the hominid-bearing Awash group, Middle Awash Valley, Afar Depression, Ethiopia. *Trans. Am. Phil. Soc.* **83**, 1–114 (1993).
60. Tassy, P. in *Geology and Palaeobiology of the Albertine Rift Valley, Uganda-Zaire, Vol. II* (eds Senut, B. & Pickford, M.) 217–257 (CIFEG, 1994).
61. Tiercelin, J.-J., Michaux, J. & Bandet, Y. Le Miocène supérieur du Sud de la dépression de l'Afar, Ethiopie: sédiments, faunes, âges isotopiques. *Bull. Soc. Geol. Fr.* **21**, 255–258 (1979).
62. Geraads, D., Alemseged, Z. & Bellon, H. The late Miocene mammalian fauna of Chorora, Awash basin, Ethiopia: systematics, biochronology and ^{40}K - ^{40}Ar age of associated volcanics. *Tertiary Res.* **21**, 113–122 (2002).
63. International Union for the Conservation of Nature. *The IUCN Red List of Threatened Species*. <http://www.iucnredlist.org/details/12392/0> (accessed 06 April 2013) (2012).
64. Roca, A. L., Georgiadis, N., Pecon-Slattery, J. & O'Brien, S. J. Genetic evidence for two species of elephant in Africa. *Science* **293**, 1473–1477 (2001).
65. Debruyne, R. *Développement Morphologique et Moléculaire des Elephantinae (Mammalia, Proboscidea)*. PhD thesis, Muséum National d'Histoire Naturelle, Paris (2003).
66. Morrison-Scott, T. C. P. A revision of our knowledge of African elephants' teeth, with notes on forest and "pygmy" elephants. *Proc. Zool. Soc. Lond.* **117**, 505–527 (1947).
67. Laws, R. M. Age criteria for the African elephant, *Loxodonta a. africana*. *East African Wildlife Journal* **4**, 1–37 (1966).
68. Aguirre, E. E. Revisión sistemática de los Elephantidae por su morfología y morfometría dentaria. *Estudios Geológicos* **24**, 109–167; **25**, 123–177, 317–367 (1969).
69. Grubb, P., Groves, C. P., Dudley, J. P. & Shoshani, J. Living African elephants belong to two species: *Loxodonta africana* (Blumenbach, 1797) and *Loxodonta cyclotis* (Matschie, 1900). *Elephant* **2**, 1–4 (2000).
70. Deino, A. L. in *Paleontology and Geology of Laetoli: Human Evolution in Context. Volume 1: Geology, Geochronology, Paleoecology and Paleoenvironment* (ed. T. Harrison) 77–98 (Springer, 2011).

Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds

Rajinder Singh^{1*}, Meilina Ong-Abdullah^{1*}, Eng-Ti Leslie Low^{1*}, Mohamad Arif Abdul Manaf¹, Rozana Rosli¹, Rajanaidu Nookiah¹, Leslie Cheng-Li Ooi¹, Siew-Eng Ooi¹, Kuang-Lim Chan¹, Mohd Amin Halim¹, Norazah Azizi¹, Jayanthi Nagappan¹, Blaire Bacher², Nathan Lakey², Steven W. Smith², Dong He², Michael Hogan², Muhammad A. Budiman², Ernest K. Lee³, Rob DeSalle³, David Kudrna⁴, Jose Luis Goicoechea⁴, Rod A. Wing⁴, Richard K. Wilson⁵, Robert S. Fulton⁵, Jared M. Ordway², Robert A. Martienssen⁶ & Ravigadevi Sambanthamurthy¹

Oil palm is the most productive oil-bearing crop. Although it is planted on only 5% of the total world vegetable oil acreage, palm oil accounts for 33% of vegetable oil and 45% of edible oil worldwide, but increased cultivation competes with dwindling rainforest reserves. We report the 1.8-gigabase (Gb) genome sequence of the African oil palm *Elaeis guineensis*, the predominant source of worldwide oil production. A total of 1.535 Gb of assembled sequence and transcriptome data from 30 tissue types were used to predict at least 34,802 genes, including oil biosynthesis genes and homologues of WRINKLED1 (*WRI1*), and other transcriptional regulators¹, which are highly expressed in the kernel. We also report the draft sequence of the South American oil palm *Elaeis oleifera*, which has the same number of chromosomes ($2n = 32$) and produces fertile interspecific hybrids with *E. guineensis*² but seems to have diverged in the New World. Segmental duplications of chromosome arms define the palaeotetraploid origin of palm trees. The oil palm sequence enables the discovery of genes for important traits as well as somaclonal epigenetic alterations that restrict the use of clones in commercial plantings³, and should therefore help to achieve sustainability for biofuels and edible oils, reducing the rainforest footprint of this tropical plantation crop.

The genus *Elaeis* (tribe *Cocoseae*) is in the family *Arecaceae*⁴, one of the oldest families of flowering plants, with fossils dating from the Cretaceous period⁵. The genus consists of two species, *E. guineensis* from West Africa⁶ and *E. oleifera* from Central and South America. *E. guineensis* has a higher yield, but *E. oleifera* has higher unsaturated fatty acid content, lower height, and resistance to disease⁷. Commercial cultivation of oil palm commenced on the West African coast in the early twentieth century⁸. In southeast Asia, where it is one of the most important commercial crops, the first recorded oil palm was brought from Africa through Mauritius and Amsterdam in 1848 (ref. 9), when four seedlings were planted as ornamentals in the Bogor Botanical Gardens in Java. Commercial cultivation began in the early twentieth century and despite the long breeding cycle (10 to 12 years) and large land requirement for field trials¹⁰, high yield breeding materials (up to 12 tonnes per hectare per year ($\text{tha}^{-1} \text{yr}^{-1}$) (ref. 9)) have been developed in less than 100 years. As such, the largely undomesticated oil palm is an ideal candidate for genomic-based tools including expressed sequence tags (ESTs)^{11–13} and transcriptome sequencing of the oil palm fruit during development, maturation and ripening^{1,14} to harness the potential of this remarkably productive crop.

We sequenced the approximately 1.8-Gb *E. guineensis* genome (AVROS (Algemene Vereniging van Rubberplanter ter Oostkust van Sumatra) *pisifera* fruit form) to high coverage with a combination of Roche/454 GS FLX Titanium (Roche/454) and Sanger bacterial artificial

chromosome (BAC) end sequencing (Supplementary Figs 1–3, Supplementary Tables 1–5 and Methods). The combined total length of the assembly (a fifth genome build or P5-build) is 1.535 Gb. Comparison of the P5-build to genetic linkage maps (Supplementary Fig. 4 and Methods) resulted in 16 genetic scaffolds (one per chromosome; 43% of the P5 scaffold assembly) for a final total of 40,072 scaffolds with an N50 (the scaffold size above which 50% of the total length of the sequence assembly can be found) of 1.27 megabases (Mb) (Supplementary Table 4) and 74% of RefSeq supported genes. For comparison, the genome of *E. oleifera* was also sequenced with a combination of fragment and linker libraries (Supplementary Table 2 and Methods). Thirty transcriptome libraries were sequenced and assembled, producing 4,528–18,936 isotigs (unique transcript assemblies) per library (Supplementary Fig. 5 and Supplementary Table 6). We sequenced 298,039 reads from methylation-filtered genomic libraries from *Deli dura* and *pisifera* genotypes of *E. guineensis* and *E. oleifera* (Methods). Methylation-filtered libraries included 90% of the gene models and were enriched 5.6 times for genes with a ‘gene space’ of between 300 and 400 Mb, comparable to that of rice and maize¹⁵.

The guanine–cytosine content of the *E. guineensis* genome (37%) is similar to that of other plant genomes, including the date palm¹⁶, but genes were conspicuous for having a much higher guanine–cytosine content (50%). Gene-finding algorithms (Methods) predicted 158,946 gene candidates covering 92 Mb of exonic gene space (5% of the 1.8-Gb genome sequence) (Supplementary Tables 4 and 5). Of these candidates, 34,802 were similar to known proteins at the peptide sequence level with 96% observed in transcriptome data (Methods). Of the remaining 124,144 candidates, 15,311 were identified in transcriptome data (Supplementary Table 4). Known retroelements (67,169) and other transposons (41,664) made up the remaining 108,833 candidates. Comparison to all repetitive element classes resulted in the identification of 775,703 independent genomic regions matching repetitive sequence elements, corresponding to 282 Mb of sequence (or 18% of the P5-build), with 39% guanine–cytosine content (Supplementary Table 4). Repeat content of the unmapped and unassembled contigs was far higher, as expected, and estimated to be approximately half of the 1.53-Gb P5-build, or 57% of the 1.8-Gb *E. guineensis* genome.

The 16 EG5 chromosomes (Fig. 1) were numbered according to size and compared with previous mapping (Supplementary Table 7) and fluorescence *in situ* hybridization studies (FISH)¹⁷. Gene density (Fig. 1a) was distributed unevenly: five of the smallest six chromosomes had one gene-rich arm, and one repeat-rich arm, as shown previously with FISH¹⁷. Known repeat classes matching the TIGR grass repeat database and REPBASE were distributed in gene-poor, methylated regions (Fig. 1b, c), whereas simple di- and trinucleotide repeats¹⁷ were mostly

¹Malaysian Palm Oil Board, 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia. ²Orion Genomics, 4041 Forest Park Avenue, St Louis, Missouri 63108, USA. ³Sackler Institute for Comparative Genomics, American Museum of Natural History, New York, New York 10024, USA. ⁴Arizona Genomics Institute, University of Arizona, Tucson, Arizona 85705, USA. ⁵The Genome Institute at Washington University, Washington University School of Medicine in St Louis, St Louis, Missouri 63108, USA. ⁶Howard Hughes Medical Institute-Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

*These authors contributed equally to this work.

within genic regions (Fig. 1d). Potential centromeric regions were identified using an internally repetitive pericentromeric repeat¹⁷ (Fig. 1e, purple), whereas highly conserved TTTAGGG telomeric repeat arrays were identified at the extreme ends of 7 of the 32 chromosome arms (Fig. 1e, green). The most prominent 5S ribosomal RNA cluster (Fig. 1e, orange) was found on the largest chromosome, whereas the only telocentric chromosome was one of the two smallest, as described previously¹⁷. Two interstitial telomere repeat arrays on chromosomes 2 and 14 were embedded within putative centromeric regions. Robertsonian fusions of telocentric chromosome ends may have given rise to these two chromosomes¹⁸, and date palm has 18 chromosomes¹⁶, consistent with this hypothesis. Typical of monocot genomes¹⁹, the most abundant repetitive elements were *copia* (33%) and *gypsy* (8%) retroelements, as well as other long terminal repeat (LTR) retrotransposons (6%) (Supplementary Fig. 6 and Supplementary Methods). Interestingly, 47% of all repeats observed were uncharacterized previously, with 73% absent from *E. oleifera* and 99% absent from *Musa acuminata* (banana). The distribution of repeats in methylation-filtered reads indicated that RIRE1 and other *copia* elements are especially heavily methylated.

Comparison of *E. guineensis* chromosomes to each other revealed that oriented homologous duplicated sequences (segmental duplications) are notably abundant (Fig. 1 and Supplementary Fig. 7). Analysis of conserved gene order revealed that the duplications were retained in *E. oleifera*, so that segmental duplications pre-dated the divergence of the African and South American oil palm (Supplementary Fig. 8a). Given that most of the genome is represented by segmental duplications, and not triplications, we conclude that oil palm is a palaeotetraploid, in line with speculation based on cytogenetics and restriction fragment length polymorphism (RFLP) mapping^{17,20}. These duplications do not span the putative pericentromeric regions (Fig. 1), indicating that most centromeres arose after polyploidization, consistent

with extensive chromosome restructuring. Homologues of 94.4%, 83.5% and 80.2% of the genes from *Phoenix dactylifera*, *M. acuminata* and *Arabidopsis* respectively, were found in *E. guineensis* (Fig. 2a, Supplementary Table 8 and Methods). Each *E. guineensis* duplication matched unique scaffolds in the date palm genome (Supplementary Fig. 8b), indicating that date palms have most of the segmental duplications found in oil palm. Polyploidization has been inferred by chromosome counts in only a limited number of *Arecaceae*²¹, and a likely scenario is that the progenitor of both palms arose as a polyploid. We performed a similar analysis of the banana genome²² and found extensive synteny between each oil palm chromosome and several chromosomes from banana (Supplementary Fig. 9), confirming that duplication events in the *M. acuminata* genome occurred after the *Musa–Elaeis* split, as proposed previously²².

The 34,802 sequence similarity gene predictions (Supplementary Table 4) were annotated for gene ontology terms (Fig. 2b, c), with a focus on oil biosynthesis (Methods). Oil synthesis in the kernel commences at 11 to 12 weeks after anthesis (WAA) and is complete by 15 to 16 WAA at which stage mesocarp oil synthesis starts, reaching a peak at 20 WAA²³. In plants, *de novo* fatty acid synthesis (FAS) is compartmentalized in plastids, whereas triacylglycerol (TAG) synthesis occurs in the cytoplasm. Although the oil palm accumulates markedly higher TAG than the date palm¹, the number of genes involved in TAG biosynthesis is notably similar in both palms. In contrast, FAS genes have higher representation in the oil palm genome (Fig. 2b). Apart from the transcriptomes of 30 tissues (Fig. 3a, b, Supplementary Fig. 5 and Supplementary Table 6), in-depth sequencing of kernel and mesocarp (Fig. 3c and Supplementary Table 9) indicated that the transcriptome signatures are similar in both tissues with plastidial FAS genes upregulated compared to TAG. Thus, the enzymes of the Kennedy pathway for TAG assembly must cope with the high flux of *de novo* fatty acids in oil palm. FAS genes were

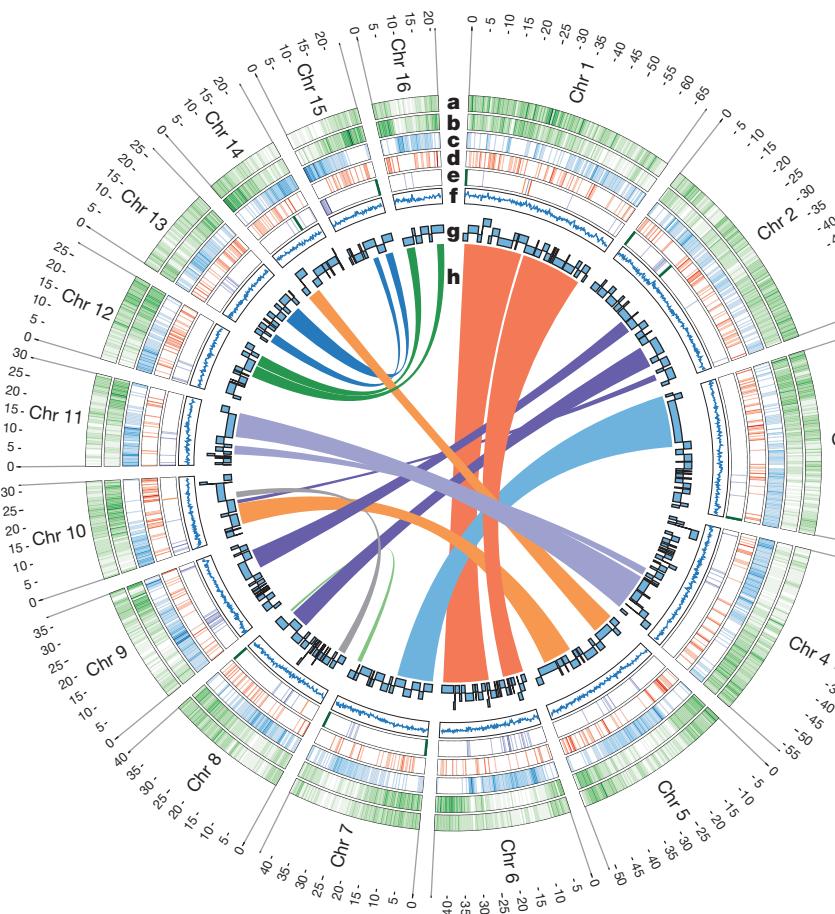


Figure 1 | The chromosomes of oil palm. *E. guineensis* has 16 chromosome pairs, ordered by size, which correspond to 16 linkage groups identified by genetic mapping (Supplementary Table 7). Tracks displayed are: a, gene density; b, methyl-filtered read density; c, retroelement density; d, simple sequence repeats; e, low copy number repetitive elements, including telomere repeat TTTAGGG (green), 5S rRNA (orange) and pericentromeric repeats (purple); f, regional G-C content (range 0.3–0.45); g, genetically mapped scaffolds from the P5-build; and h, segmental duplications. Densities for telomere repeats are exaggerated for visual clarity.

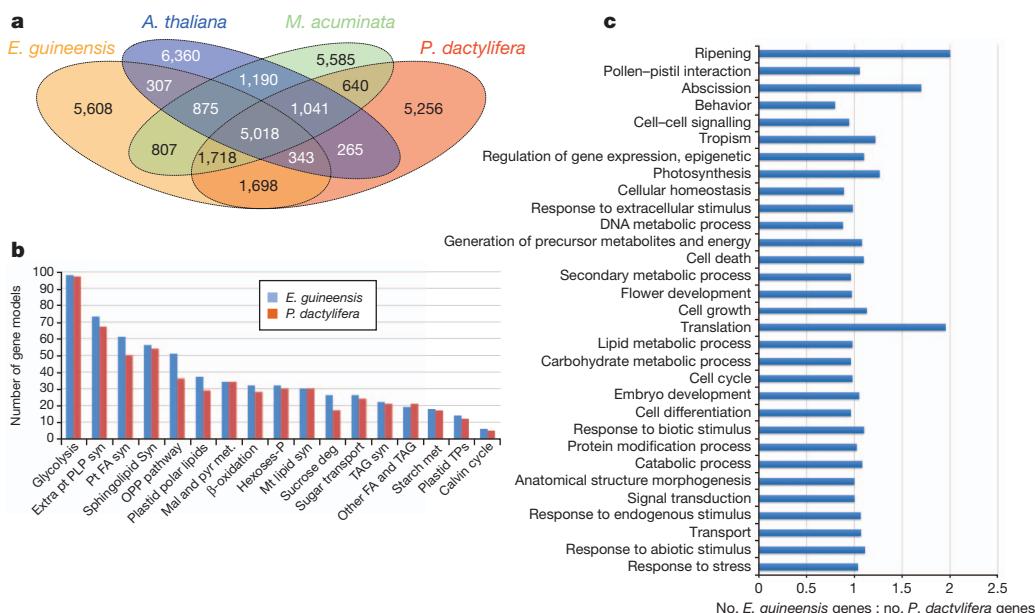


Figure 2 | Gene model comparisons. **a**, Venn Diagram illustrating the proportion of shared gene family clusters in *E. guineensis*, *M. acuminata*, *A. thaliana* and *P. dactylifera*. Number of genes (clusters) compared were: *A. thaliana*, 27,416 (15,399); *M. acuminata*, 36,529 (16,874); *P. dactylifera*, 28,882 (15,979) and *E. guineensis*, 34,802 (16,374). **b**, Gene ontology

classifications of oil palm and date palm. **c**, Ratio of gene number (oil palm:date palm) in each gene ontology classification. Deg, degradation; FA, fatty acid; hexoses-P, hexose phosphate pathway; mal and pyr met, malate and pyruvate metabolism; mt, mitochondrial; OPP, oxidative pentose phosphate; pt PLP syn, plastidial phospholipid synthesis; TAG, triacylglycerol; TPs, transporters.

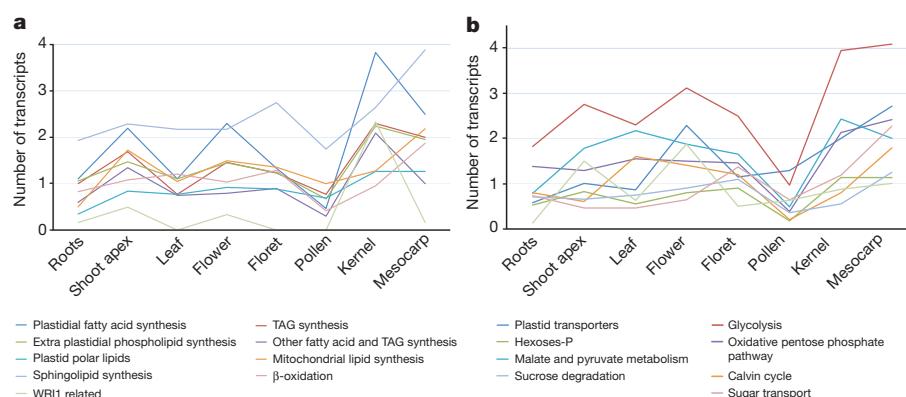
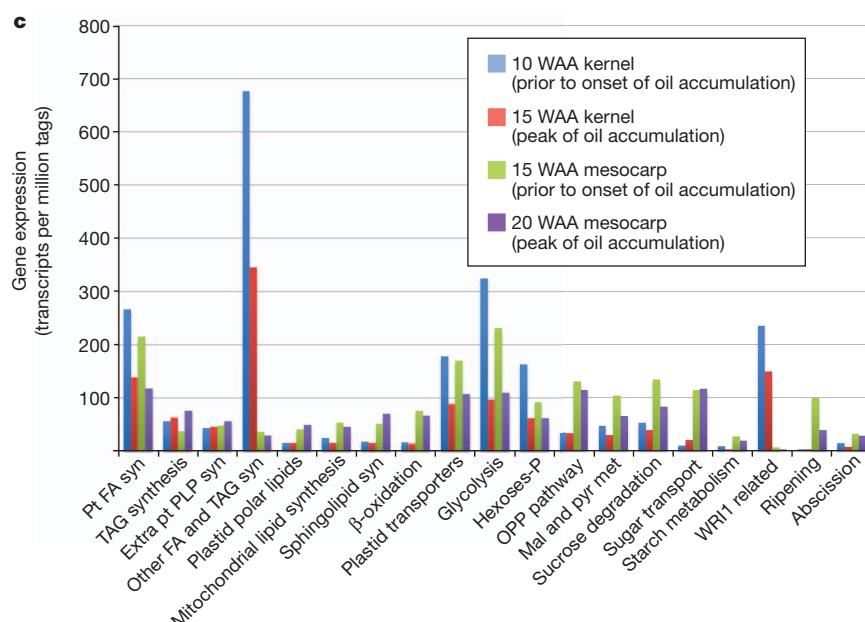


Figure 3 | Lipid and carbohydrate metabolism in oil palm fruits. **a, b**, Number of lipid-synthesis-related (a) and carbohydrate-synthesis-related (b) gene transcripts in different tissues. **c**, Comparison of gene-expression levels between kernel and mesocarp tissue before and at peak of oil accumulation.



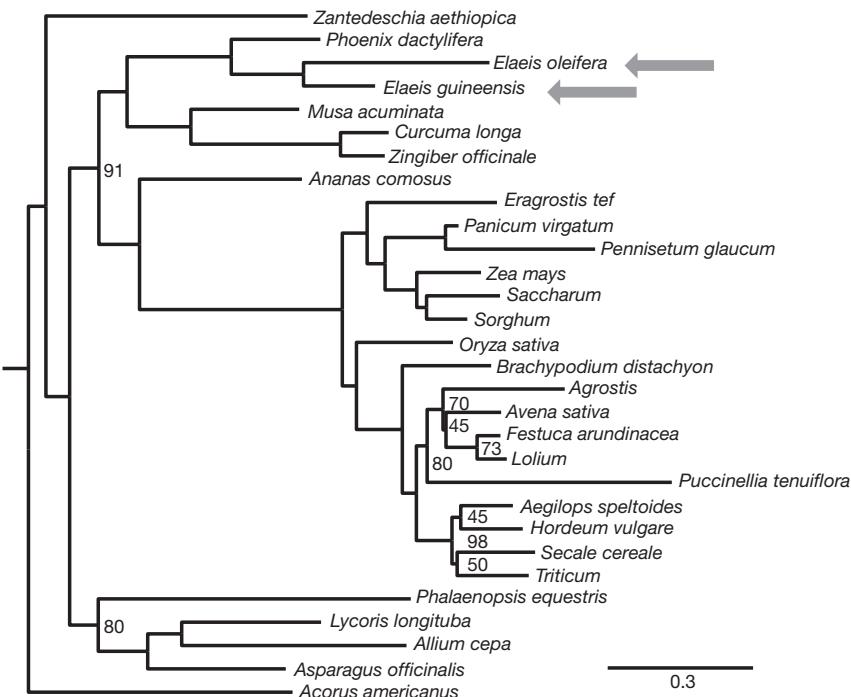


Figure 4 | Phylogenetic analysis. A carefully annotated subset of proteins from *E. guineensis*, *E. oleifera* and *P. dactylifera* were included in a matrix of 1,685 gene partitions (858,954 patterns) and 107 taxa. This matrix is extracted from partitions with at least 30 taxa present in a much larger matrix.

upregulated markedly just before the onset, and then declined during the peak of lipid accumulation. This contrasts with previous reports that FAS transcripts continue to increase in the mesocarp¹. This may reflect the *tenera* fruit form used in this study, as *dura* has a longer maturation period (22 WAA)²⁴ than *tenera* (20 WAA)^{23,25}. WRI1 regulates oil accumulation in the oil palm mesocarp¹ and we found highest mesocarp expression of WRI1 just before lipid accumulation onset. However, kernel showed 75-fold higher WRI1 expression compared to mesocarp (Fig. 3c and Supplementary Table 9), implying its pivotal role in kernel oil synthesis. *LEAFY COTYLEDON1* (*LEC1*), *LEC2*, *ABSCISIC ACID INSENSITIVE3* (*ABI3*) and *FUSCA3*, which activate *WRI1* in oilseeds, were not found in mesocarp transcriptomes, but were well-represented in the kernel, with *LEC1* and *ABI3* showing higher expression at the start and completion of oil accumulation, respectively. Interestingly, the transcriptional regulator *PICKLE* (*PKL*) was expressed in both the mesocarp and kernel (Supplementary Table 9).

Genes involved in sucrose degradation and the oxidative pentose phosphate pathway were more highly represented in oil palm than date palm (Fig. 2b). Pentose phosphates are recycled into glucose 6-phosphate to fuel glycolytic pathways, and import of these cytosolic metabolites requires specific transporters on the plastid envelope. Although these transporter genes are upregulated strongly in oil palm¹, they are similarly represented in date and oil palm genomes (Fig. 2b). Thus, channelling of sugars destined for oil synthesis is regulated at the transcriptome level in oil palm. Additional insights important to TAG biosynthesis, fruit ripening and abscission are provided in Supplementary Notes.

To place palms on the evolutionary tree, evidence-based gene models from each species were combined with a previous seed plant data set²⁶ to form a matrix of 1,685 gene partitions (858,954 patterns) and 107 taxa. *P. dactylifera*, *E. guineensis* and *E. oleifera* are present in 1,206, 1,229 and 1,190 partitions, respectively. All three were well separated from other monocots (Fig. 4), including nearest neighbours *Musa* (banana), *Curcuma* (turmeric) and *Zingiber* (ginger). Phylogenetic dating using conservative constraints (Methods) predicted a divergence 65 million

years (Myr) ago between date and oil palm, and 51 Myr between *E. oleifera* and *E. guineensis*. This is comparable with divergence between Old and New world relatives such as African *Sorghum bicolor* (sorghum) and American *Zea mays* (maize) panicoid grasses (26 Myr). Unlike maize and sorghum, however, *E. guineensis* and *E. oleifera* give rise to fertile hybrids², consistent with the vicariant hypothesis for phylogeographical divergence, in which geographically isolated species are under no selective pressure to evolve reproductive isolation²⁷.

The genome sequence of oil palm will be a rich resource for oil palm breeders, geneticists and evolutionary biologists alike. It has revealed that palms are ancient tetraploids, and that the African and South American species probably diverged in the Old and New Worlds. Overrepresented genes in lipid and carbohydrate metabolism are expressed differentially in mesocarp and kernel, accounting for the different properties of palm fruit and palm kernel oils. The genome sequence will also enable mapping of somaclonal epigenetic alterations that restrict the use of clones in commercial plantings. The dense representation of sequenced scaffolds on the genetic map will facilitate identification of genes responsible for important yield and quality traits. The genome sequence of this tropical plantation crop is an important step in achieving the goals that are critical to the sustainability challenges associated with growing demands for biofuels and edible oils.

METHODS SUMMARY

Genome sequencing and assembly, genetic mapping and gene annotation. Sequencing reads (Roche/454) were generated from genomic and BAC pool DNA fragment and paired-end linker libraries (Supplementary Figs 1–3 and Supplementary Tables 1–3). BAC end sequencing was performed using the Sanger method. Sequence data were assembled as described (Methods). The genetic map for the selfed Nigerian *tenera* palm T128 was constructed as described (Methods). Gene predictions, genome and protein comparisons and gene ontology annotation were performed as described (Methods).

Transcriptome sequencing. Thirty transcriptome libraries were constructed and transcriptome sequences were generated (Roche/454) (Methods). Transcriptome libraries from kernel and mesocarp were constructed and deep sequenced by Illumina HiSeq 2000 (Methods).

Methylation filtration. Methylation filtration libraries were constructed as described¹⁵ and resulting clones were sequenced using the Sanger method (Methods). **Phylogenetic analysis.** Open reading frames were assembled from evidence-based gene models and combined with a large seed plant data set²⁶ (Methods).

Full Methods and any associated references are available in the online version of the paper.

Received 30 September 2012; accepted 16 May 2013.

Published online 24 July; corrected online 14 August 2013 (see full-text HTML version for details).

- Bourgis, F. et al. Comparative transcriptome and metabolite analysis of oil palm and date palm mesocarp that differ dramatically in carbon partitioning. *Proc. Natl Acad. Sci. USA* **108**, 12527–12532 (2011).
- Hardon, J. J. & Tan, G. Y. Interspecific hybrids in the genus *Elaeis* I. crossability, cytogenetics and fertility of F₁ hybrids of *E. guineensis* × *E. oleifera*. *Euphytica* **18**, 372–380 (1969).
- Jaligot, E. et al. Epigenetic imbalance and the floral developmental abnormality of the *in vitro*-regenerated oil palm *Elaeis guineensis*. *Ann. Bot.* **108**, 1453–1462 (2011).
- Dransfield, J. et al. *Genera Palmarum: The Evolution and Classification of Palms* (Royal Botanic Gardens Kew, 2008).
- Purseglove, J. W. In *Tropical Crops (Monocotyledons)* 416–510 (Longman, 1972).
- Zeven, A. C. The origin of the oil palm. *J. Niger. Inst. Oil Palm Res.* **4**, 218–225 (1965).
- Cochard, B., Amblard, P. & Durand-Gasselin, T. Oil palm genetic improvement and sustainable development. *Oleagineux Corps Gras Lipides* **12**, 141–147 (2005).
- Hartley, C. *The Oil Palm* 47–94 (Longman, 1988).
- Corley, R. H. V. & Tinker, P. B. In *The Oil Palm* 4th edn, 1–26 (Blackwell Science, 2003).
- Mayes, S., Jack, P. L., Corley, R. H. & Marshall, D. F. Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* **40**, 116–122 (1997).
- Jouannic, S. et al. Analysis of expressed sequence tags from oil palm (*Elaeis guineensis*). *FEBS Lett.* **579**, 2709–2714 (2005).
- Ho, C.-L. et al. Analysis and functional annotation of expressed sequence tags (ESTs) from multiple tissues of oil palm (*Elaeis guineensis* Jacq.). *BMC Genomics* **8**, 381 (2007).
- Low, E. T. L. et al. Oil palm (*Elaeis guineensis* Jacq.) tissue culture ESTs: identifying genes associated with calllogenesis and embryogenesis. *BMC Plant Biol.* **8**, 62 (2008).
- Tranbarger, T. J. et al. Regulatory mechanisms underlying oil palm fruit mesocarp maturation, ripening, and functional specialization in lipid and carotenoid metabolism. *Plant Physiol.* **156**, 564–584 (2011).
- Palmer, L. E. et al. Maize genome sequencing by methylation filtration. *Science* **302**, 2115–2117 (2003).
- Al-Dous, E. K. et al. De novo genome sequencing and comparative genomics of date palm (*Phoenix dactylifera*). *Nature Biotechnol.* **29**, 521–527 (2011).
- Castilho, A. M., Vershinin, A. V. & Heslop-Harrison, J. S. Repetitive DNA and the chromosomes in the genome of oil palm (*Elaeis guineensis*). *Ann. Bot.* **85**, 837–844 (2000).
- Richards, E. J., Chao, S., Vongs, A. & Yang, J. Characterization of *Arabidopsis thaliana* telomeres isolated in yeast. *Nucleic Acids Res.* **20**, 4039–4046 (1992).
- Du, J. et al. Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* **63**, 584–598 (2010).
- Singh, R. et al. Identification of cDNA RFLP markers and their use for molecular mapping in oil palm. *Asia Pac. J. Mol. Biol. Biotechnol.* **16**, 53–63 (2008).
- Wood, T. E. et al. The frequency of polyploid speciation in vascular plants. *Proc. Natl Acad. Sci. USA* **106**, 13875–13879 (2009).
- D'Hont, A. et al. The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. *Nature* **488**, 213–217 (2012).
- Sambanthamurthi, R., Sundram, K. & Tan, Y. Chemistry and biochemistry of palm oil. *Prog. Lipid Res.* **39**, 507–558 (2000).
- Bafor, M. E. & Osagie, A. U. Changes in lipid class and fatty acid composition during maturation of mesocarp of oil palm (*Elaeis guineensis*) variety dura. *J. Sci. Food Agric.* **37**, 825–832 (1986).
- Shaarani, S. M., Cárdenas-Blanco, A., Amin, M. H. G., Soon, N. G. & Hall, L. D. Monitoring development and ripeness of oil palm fruit (*Elaeis guineensis*) by MRI and bulk NMR. *Int. J. Agric. Biol.* **12**, 101–105 (2010).
- Lee, E. K. et al. A functional phylogenomic view of the seed plants. *PLoS Genet.* **7**, e1002411 (2011).
- Riggins, C. W. & Seigler, D. S. The genus *Artemisia* (Asteraceae: Anthemideae) at a continental crossroads: molecular insights into migrations, disjunctions, and reticulations among Old and New World species from a Beringian perspective. *Mol. Phylogenet. Evol.* **64**, 471–490 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge the contributions of N. Ahmad, M. Marjuni and N. Abdullah for sampling of oil palm materials. We thank MoGene, GeneWorks, Beijing Genome Institute, Tufts University Core Facility and Macrogen for sequencing services. We thank D. Stevenson for discussions. We appreciate the constant support of Y. M. Choo, and the Ministry of Plantation Industries and Commodities, Malaysia. The project was endorsed by the Cabinet Committee on the Competitiveness of the Palm Oil Industry (CCPO) and funded by the Malaysian Palm Oil Board. E.K.L., R.D. and R.A.M. are supported by a grant from NSF 0421604 (Genomics of Comparative Seed Plant Evolution).

Author Contributions R.S., M.O.-A., E.-T.L.L. and R.S. conceptualized the research programme. R.S., M.O.-A., E.-T.L.L., R.N., N.L., J.M.O., S.W.S., R.K.W., R.S.F., R.A.M. and R.S. designed experiments and coordinated the project. R.S., M.O.A., E.-T.L.L., M.A.A.M., L.C.-L.O., S.-E.O., J.N., B.B., M.A.B., S.W.S., J.M.O. and R.S. conducted laboratory experiments and were involved in data analysis. E.-T.L.L., R.R., K.-L.C., M.A.H., N.A., S.W.S., D.H. and M.H. performed bioinformatics analyses. E.K.L. and R.D. carried out phylogenetic analysis. D.K., J.L.G. and R.A.W. built BAC libraries and carried out BAC end sequencing. R.S., M.O.A., E.-T.L.L., R.N., N.L., S.W.S., J.M.O., R.A.M. and R.S. participated in preparing and revising the manuscript. R.A.M. and R.S. supervised data generation and analysis.

Author Information The *E. guineensis* and *E. oleifera* BioProjects are available for download at <http://genomsawit.mpopb.gov.my> and have been registered at the NCBI under BioProject accessions PRJNA192219 and PRJNA183707. The Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accessions ASJS00000000 (*E. guineensis*) and ASIR00000000 (*E. oleifera*). The versions described in this paper are versions ASJS00000000 and ASIR00000000. Reprints and permissions information are available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.A.M. (martiens@cshl.edu) or R.S. (raviga@mpob.gov.my).

 This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported licence. To view a copy of this licence, visit <http://creativecommons.org/licenses/by-nc-sa/3.0>

METHODS

Genome assembly. The assembly of *E. guineensis* (AVROS, *pisifera* fruit form) genome P5-build was constructed from sequences from a total of 148 linker libraries and 81 fragment libraries (Roche/454). Reads were generated from genomic DNA fragment libraries (53.5 million reads), BAC pool DNA fragment libraries (3.6 million reads), and from a series of genomic (89.1 million reads) and BAC (8.6 million reads) paired-end linker libraries. In total, 46.8 billion bases of raw sequence were generated, representing approximately 26-fold raw sequence coverage of the 1.8-Gb oil palm genome. Sequence data were assembled using the Newbler algorithm²⁸ (Supplementary Table 4). The seventh assembly and genome build of *E. oleifera* (O7-build) was constructed from a total of 127 linker libraries and 68 fragment libraries. In total, 130 million *E. oleifera* reads, representing approximately 25-fold raw sequence coverage, were generated (Supplementary Table 2).

High information content restriction fragment fingerprints were generated for 124,286 BAC clones from the *E. guineensis* genome with an average size of 150 kb. BAC fingerprints were used to construct a physical map of the reference genome (Supplementary Figs 2 and 3). In addition, BAC ends from this library were sequenced using Sanger sequencing to generate 235,613 paired reads with an average read length of approximately 600 bp. BAC end reads were used to create the shotgun assembly, as well as to locate individual BACs within the assembled genome.

Before assembly, all Roche/454 sequence runs were screened for quality based on average read length, linker library efficiency and library redundancy using a custom pipeline based on the SSAHA program²⁹. For linker positive reads, library insert sizes were estimated by aligning both ends to a draft assembly of oil palm, and measuring the intervening sequence in cases in which both ends matched a single draft scaffold. In order to minimize the negative impact of chimaeras in the linker libraries, we removed all identical reads that were due to library redundancy, not independent observations. In addition to the Roche/454 data, a set of 235,613 BAC end reads from the Origen_1 BAC library were included in the P5 assembly, with an average BAC size of 150 kb. These data were assembled using the Newbler algorithm²⁸ on a Dell PowerEdge R910 server with 512 Gb of RAM, and 32 cores running Ubuntu Linux 10.04. The P5 assembly took 15 days and 5 h to complete.

Genetic map construction. Two genetic maps were constructed. The first mapping population was derived from the self-pollination of the Nigerian *tenera* palm T128. The mapping population and map-construction methodology are described elsewhere³⁰. The second mapping population comprised 87 palms obtained from a cross between Ulu Remis Deli *dura* (ENL48) and Yangambi *pisifera* (ML161) grown at the FELDA Research Station at Jerantut, Malaysia. Linkage analysis for the *dura* × *pisifera* (P2) cross was performed using both JoinMap 4.0 and GenStat 14th edition. JoinMap was used to examine markers and identify loci showing distorted segregation (chi-squared test). JoinMap was initially used to construct the two parental framework maps at recombination frequency ≤ 0.2 and a nearest neighbour stress value of ≤ 4 (centimorgan) using the maximum likelihood mapping algorithm as described³¹. The density of the linkage maps was later increased by mapping additional co-dominant markers into the parental framework maps using the maximum likelihood mapping algorithm in GenStat 14. The integrated map was built using the maximum likelihood mapping method in GenStat14 by combining data from markers on both the two parental maps. Comparison between the integrated map and the parental maps was visualized using MapChart 2.2 (ref. 32).

Genetic-map integration and chromosome-sequence construction. The 1,511 markers used in the generation of the T128_codominant and P2_DxP maps were compared to scaffolds from the P5-build using the exonerate³³ program with an ungapped alignment and a minimum identity match of 97%. Markers that did not uniquely map to P5 scaffolds were discarded, and one scaffold marker ordering was created for each of the two maps. After reviewing shared scaffolds between the two maps, a final order was determined for ordering 169 scaffolds, and ordering and orienting 124 scaffolds based on multiple markers. The scaffold sequences were then concatenated in order and reverse complemented as required to create 16 linkage groups based chromosome sequences. During map integration, LG15 (T128_codominant naming convention) appeared significantly shorter than a previous integration based on the P4 assembly with the P2_regression map. This was owing to map instability introduced into the P2_DxP map generation in which the P2_regression was more stable. After review, the LG15 chromosome sequence was extended based on mapping of P5 scaffolds onto the P4/P2_regression version. **Gene identification and annotation.** Based on the longest set of scaffolds representing 10% of the original P1-build, we used the SNAP³⁴ gene finder to identify initial candidate gene predictions for the *E. guineensis* genome. Initial SNAP runs were performed using the rice (*O.sativa*) gene model. Initial genes discovered were compared with the RefSeq35 database as well as the TIGR Gramineae repeat database in order to remove retroelements and pseudogenes. The remaining

transcripts were then screened for missing start and stop codons, as well as other warnings from SNAP. The remaining set was then used as input to the SNAP programs FATHOM and FORGE according to the SNAP documentation to create a new hidden Markov model (HMM) with greater specificity to *E. guineensis*. The same screening process was applied again and four iterations of the training were used resulting in the final ‘pisif_2_22_11.hmm’ gene model.

For the Glimmer³⁵ predictions, assembled *E. guineensis* transcriptome sequences (groups A, C, D and G in Supplementary Table 6) were translated from start to stop with a size selection ranging from 500 to 5,000 nucleotides. These were then compared to complete coding sequences for Magnoliophyta from GenBank using BLASTX (*E*-value < 10⁻¹⁰). Transcripts with homology starting at position one of the Magnoliophyta targets were selected for further analysis. The transcripts were then screened using BLASTClust (NCBI) and CD-HIT³⁶ to reduce the number of genes to meet Glimmer’s training requirements. Exon boundaries were determined by mapping to the previous P4-build, and were used to create an *E. guineensis* HMM.

Transcriptome analysis. Reads (Roche/454) from each library and from all libraries (Fig. 3a, b and Supplementary Table 6) were assembled into isotigs respectively. The isotigs from each library were blasted on *A. thaliana* gene models with a threshold of *E*-value < 10⁻⁵. The best hit *A. thaliana* gene model was assigned to the homologue of the query isotig. Based on Bourgis’ annotation¹, copy numbers were given to each gene in each category group. The final copy numbers of each functional group were scaled on the number of genes in each group. To estimate expression level of genes in mesocarp and kernel tissues, Illumina HiSeq 2000 reads from each library were mapped to assembled isotigs from all *E. guineensis* reads by using the Burrows–Wheeler Aligner. Gene group expression levels were calculated as the number of mapped reads on each isotig divided by the total number of isotigs, multiplied by 100,000, and scaled by the number of genes in each gene group. Both copy number and read coverage were the mean of measures from two biological replicates. Data were analysed as described above for Roche/454 data, except that expression levels were calculated as transcripts per million tags.

Methylation-filtered library analysis. Methylation-filtered (“GeneThresher”) genomic DNA libraries were constructed as described^{15,37} to select unmethylated clones (depleted of most repetitive sequences³⁸) by propagation in McrBC⁺ strains of *Escherichia coli*. Briefly, nuclear DNA was extracted individually from young leaves of Deli *dura* and AVROS *pisifera* of *E. guineensis* and from *E. oleifera*. For each of the three DNA populations, genomic shotgun libraries were constructed as described³⁷. Sequences were generated from one end of each cloned insert by ABI 3730 sequencing (Life Technologies), generating 298,039 reads (73,390 from Deli *dura*, 101,327 from AVROS *pisifera* and 123,322 from *E. oleifera*).

Segmental duplication analysis. Chromosomes from the EG5-linked assembly were screened in a self-self comparison test using the MUMmer3 set of tools³⁹. Final alignments were carried out on chromosome pairs using the PROmer program (optional parameters –d 0.5 –c 200), alignments were reviewed visually with the MUMmer plot program, and approximate boundaries for segmental duplications were recorded (Supplementary Fig. 7 and Supplementary Table 7).

To test for the existence of observed segmental duplications in other genomes, we performed comparative genomics of each half of the 16 segmental duplications in *E. guineensis* with the *E. oleifera* and *P. dactylifera* scaffold sets. Comparisons were performed using a custom analysis pipeline based on the MUMmer program (optional parameters –n –l 30 –b –c –L). MUMmer output was summarized as an offset-sorted overlap plot showing where query scaffolds share local alignment with a reference chromosome. Output was reviewed to verify that each half of the proposed segmental duplication was present in the query genome, and that the scaffolds matching were different for each half of the duplication in *E. guineensis*. **Genome comparison by gene models.** NCBI TBLASTN program was used to compare *A. thaliana*, *O. sativa*, *P. dactylifera* and *E. guineensis* predicted proteins on the genomes of each of the four species, with a threshold of *E*-value < 10⁻⁵. At this level of conservation, matches represent shared gene families between the query and target genomes. By comparing predicted proteins to genome sequence, biases introduced by gene prediction methods are minimized compared to a direct gene-level comparison. Comparisons between the gene models from one species and its own genome are close to, but less than 100%, owing to the limit of sensitivity of TBLASTN at this *E*-value cutoff. Public database sources for comparative genome sequences and gene models are provided in Supplementary Methods.

Gene clustering and Venn diagram. CD-HIT clustering algorithm³⁶ was used to look for homologous protein sequences among *A. thaliana*, *O. sativa*, *P. dactylifera* and *E. guineensis* at 40% similarity level. This algorithm avoids all-versus-all BLAST search by using a short word filter.

Gene ontology annotation. *E. guineensis* gene model protein sequences were queried against *A. thaliana* annotated gene model protein sequences by using BLASTP⁴⁰ with a threshold of *E*-value < 10⁻⁵. The blast output file was then loaded into Blast2GO⁴¹. Blast2GO performed gene ontology annotation by using

an annotation rule on the found ontology terms. The most specific annotations were assigned on each sequence with default parameters in the annotation rule.

Phylogenetic analysis. A carefully annotated set of open reading frames was assembled from evidence-based gene models in each species. The oil palm open reading frames were combined with a large seed plant data set published previously²⁶, with updated gene models for *Arabidopsis thaliana*, *Oryza sativa* and *Solanum lycopersicum*, as well as the addition of *P. dactylifera*, *M. acuminata*, *Carica papaya* and *Selaginella moellendorffii* (outgroup) for a total of 107 taxa. Orthologues were identified using OrthoLogID⁴², and a sub-matrix with at least 30 taxa present per partition was extracted for phylogenetic analysis. This sub-matrix contains 1,685 gene partitions with 858,954 patterns. Maximum likelihood analysis was carried out using RAxML⁴³ with the JTT+F+Γ model and 100 bootstrap replicates.

28. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. *Nature* **437**, 376–380 (2005).
29. Ning, Z., Cox, A. J. & Mullikin, J. C. SSAHA: a fast search method for large DNA databases. *Genome Res.* **11**, 1725–1729 (2001).
30. Singh, R. et al. The oil palm *SHELL* gene controls oil yield and encodes a homologue of SEEDSTICK. *Nature* <http://dx.doi.org/10.1038/nature12356> (this issue).
31. Jansen, J. Construction of linkage maps in full-sib families of diploid outbreeding species by minimizing the number of recombinations in hidden inheritance vectors. *Genetics* **170**, 2013–2025 (2005).
32. Voorrips, R. E. MapChart: software for the graphical presentation of linkage maps and QTLs. *J. Hered.* **93**, 77–78 (2002).
33. Slater, G. S. & Birney, E. Automated generation of heuristics for biological sequence comparison. *BMC Bioinformatics* **6**, 31 (2005).
34. Korf, I. Gene finding in novel genomes. *BMC Bioinformatics* **5**, 59 (2004).
35. Majoros, W. H., Pertea, M. & Salzberg, S. L. TigrScan and GlimmerHMM: two open source ab initio eukaryotic gene-finders. *Bioinformatics* **20**, 2878–2879 (2004).
36. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
37. Bedell, J. A. et al. Sorghum genome sequencing by methylation filtration. *PLoS Biol.* **3**, e13 (2005).
38. Ouyang, S. & Buell, C. R. The TIGR Plant Repeat Databases: a collective resource for the identification of repetitive sequences in plants. *Nucleic Acids Res.* **32**, D360–D363 (2004).
39. Kurtz, S. et al. Versatile and open software for comparing large genomes. *Genome Biol.* **5**, R12 (2004).
40. Altschul, S. F. et al. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
41. Conesa, A. et al. Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* **21**, 3674–3676 (2005).
42. Chiu, J. C. et al. OrthoLogID: automation of genome-scale ortholog identification within a parsimony framework. *Bioinformatics* **22**, 699–707 (2006).
43. Stamatakis, A. RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. *Bioinformatics* **22**, 2688–2690 (2006).

The oil palm *SHELL* gene controls oil yield and encodes a homologue of *SEEDSTICK*

Rajinder Singh¹, Eng-Ti Leslie Low¹, Leslie Cheng-Li Ooi¹, Meilina Ong-Abdullah¹, Ngoot-Chin Ting¹, Jayanthi Nagappan¹, Rajanaidu Nookiah¹, Mohd Din Amiruddin¹, Rozana Rosli¹, Mohamad Arif Abdul Manaf¹, Kuang-Lim Chan¹, Mohd Amin Halim¹, Norazah Azizi¹, Nathan Lakey², Steven W. Smith², Muhammad A. Budiman², Michael Hogan², Blaire Bacher², Andrew Van Brunt², Chunyan Wang², Jared M. Ordway², Ravigadevi Sambanthamurthi¹ & Robert A. Martienssen³

A key event in the domestication and breeding of the oil palm *Elaeis guineensis* was loss of the thick coconut-like shell surrounding the kernel. Modern *E. guineensis* has three fruit forms, *dura* (thick-shelled), *pisifera* (shell-less) and *tenera* (thin-shelled), a hybrid between *dura* and *pisifera*^{1–4}. The *pisifera* palm is usually female-sterile. The *tenera* palm yields far more oil than *dura*, and is the basis for commercial palm oil production in all of southeast Asia⁵. Here we describe the mapping and identification of the *SHELL* gene responsible for the different fruit forms. Using homozygosity mapping by sequencing, we found two independent mutations in the DNA-binding domain of a homologue of the MADS-box gene *SEEDSTICK* (*STK*, also known as *AGAMOUS-LIKE 11*), which controls ovule identity and seed development in *Arabidopsis*. The *SHELL* gene is responsible for the *tenera* phenotype in both cultivated and wild palms from sub-Saharan Africa, and our findings provide a genetic explanation for the single gene hybrid vigour (or heterosis) attributed to *SHELL*, via heterodimerization. This gene mutation explains the single most important economic trait in oil palm, and has implications for the competing interests of global edible oil production, biofuels and rainforest conservation⁶.

Oil palm fruits (drupes) are derived from three fused carpels and consist of epicarp, mesocarp and endocarp tissues surrounding one or more kernels. Hybrids (*tenera*) between *dura* and *pisifera* (Fig. 1) have a distinct fibre ring that surrounds the coconut-like shell of the oil palm seed⁷. The *SHELL* gene responsible for this phenotype has co-dominant monogenic inheritance, first reported in the Belgian Congo in the 1940s (ref. 7). However, *tenera* fruit forms were recognized and exploited in Africa well before then^{2,8}. Given the central role played by the *SHELL* gene, oil palm breeding uses reciprocal recurrent selection of maternal (*dura*) and paternal (*pisifera*) pools⁴. The Deli *dura* population, direct descendants of the four original African palms planted in the Bogor Botanical Garden, Indonesia (1848), has excellent combining ability with the AVROS (Algemene Vereniging van Rubberplanters ter Oostkust van Sumatra) *pisifera* and other *pisifera* parental palms. AVROS *pisifera* palms were derived from the famous ‘Djongo’ palm from Congo, but more recently several different accessions of *dura* and *tenera* have also been sourced from Africa⁴. *Tenera* palms are thought to have been selected by pre-colonial cultures in West Africa owing to their higher oil yields, and are the basis for modern oil palm breeding⁴.

The *SHELL* gene lies 4.7 and 9.8 cM away from the closest molecular markers^{9–11}, but has proven extremely challenging to identify given the large genome, long generation times and difficulty of phenotyping in experimental populations of oil palm, which are widely distributed among different plantations (Methods). We used a two-tiered approach to identify the *SHELL* gene, taking advantage of the recently completed oil palm genome sequence¹². First, 240 F₁ progeny, derived by controlled self-pollination of the Nigerian *tenera* accession T128 and grown over two decades in plantations throughout Malaysia, were

scored for fruit form phenotype¹³. In addition to 200 restriction fragment length polymorphism (RFLP) and simple sequence repeat (SSR) markers, the progeny were genotyped for 4,451 single nucleotide polymorphism (SNP) markers derived from the oil palm genome sequence¹² by the Infinium iSelect Assay (Illumina). A genetic map with 16 linkage groups was constructed, and the *SHELL* gene locus was placed in T128 linkage group 7 (Supplementary Fig. 1), consistent with previous studies (Methods), and mapped by sequence similarity to a 3.4-megabase (Mb) assembly scaffold 43 (p3-sc00043) at the end of chromosome 2 (ref. 12). A tiling path of bacterial artificial chromosome (BAC) contigs corresponding to scaffold 43 was selected from a high-information content physical map of *pisifera* and sequenced¹². Further SNP assays were designed from an improved assembly, and additional

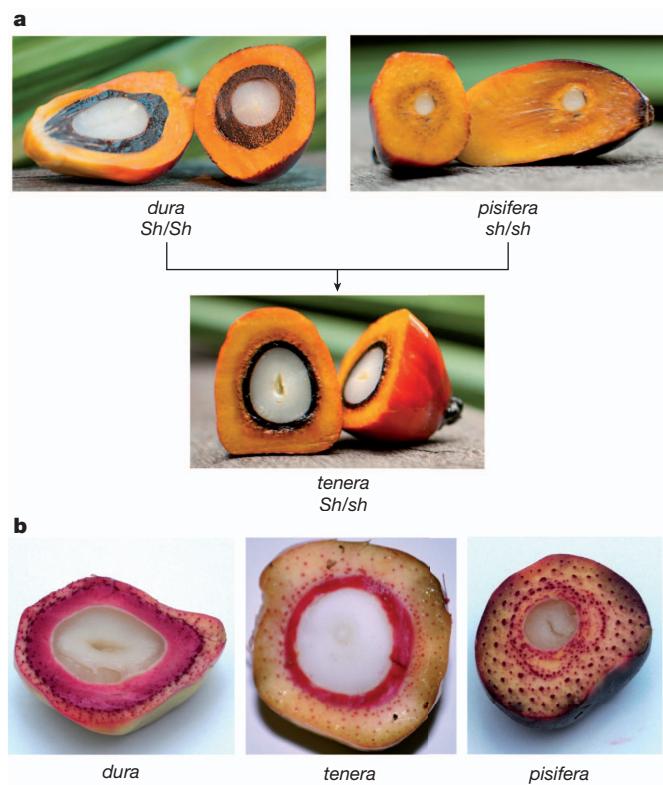


Figure 1 | Fruit forms of the African oil palm *Elaeis guineensis*. a, *Dura* (*Sh/Sh*) fruit forms have a thick lignified shell surrounding the kernel, which is absent in *pisifera* (*sh/sh*). F₁ hybrid palms (*Sh/sh*) have an intermediate fruit form (*tenera*) that is much higher yielding than either parent in terms of mesocarp oil. **b**, *Dura*, *tenera* and *pisifera* fruit stained for lignin (red) with phloroglucinol.

¹Malaysian Palm Oil Board, 6, Persiaran Institusi, Bandar Baru Bangi, 43000 Kajang, Selangor, Malaysia. ²Orion Genomics, 4041 Forest Park Avenue, St Louis, Missouri 63108, USA. ³Howard Hughes Medical Institute–Gordon and Betty Moore Foundation, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York 11724, USA.

genotypes incorporated into the map. In a review of single marker data, recombinant breakpoints were identified, indicating that the gene lay in a 450-kilobase (kb) interval (Methods).

Next, we used homozygosity mapping using the AVROS pedigree (Fig. 2a) and whole-genome resequencing (Methods). In this technique, candidate genes appear as regions with low diversity in homozygous inbred individuals^{14,15}. There were 14 individual *pisifera* palm genomes sequenced at 20× genomic coverage, and 29 additional *pisifera* palms sequenced as a pool (pool 1) at 40× coverage (Illumina HiSeq 2000). SNPs were called throughout the genome, and those in scaffold 43 were scored for homozygosity. The resulting homozygosity plot had a local minimum of 200 kb (centred on 350,000 base pairs (bp) in p3-sc00043; Fig. 2b, c). This 200-kb region contained about 30 annotated genes,

only five of which were fully homozygous, and only one of which lay in the genetic interval containing *SHELL* (Fig. 2c). This gene encodes a homologue of *SEEDSTICK*, which is responsible for ovule and seed development in *Arabidopsis*^{16,17} (Fig. 2c and Supplementary Fig. 2).

PCR amplicon sequencing (Methods) identified allelic differences between *SHELL* in *Deli dura* (*Sh*^{DeliDura}), and in the AVROS (*sh*^{AVROS}) and T128 (*sh*^{MPOB}) *pisifera* haplotypes derived from Congo and Nigeria, respectively (Fig. 3). A nucleotide substitution in *sh*^{MPOB} results in a leucine-to-proline amino acid change in the conserved DNA binding and dimerization domain, whereas a substitution in *sh*^{AVROS} results in a lysine-to-asparagine amino acid change in the same domain, only two amino acids carboxy-terminal to the *sh*^{MPOB} change (Fig. 3). In related proteins, this highly conserved lysine residue is involved in

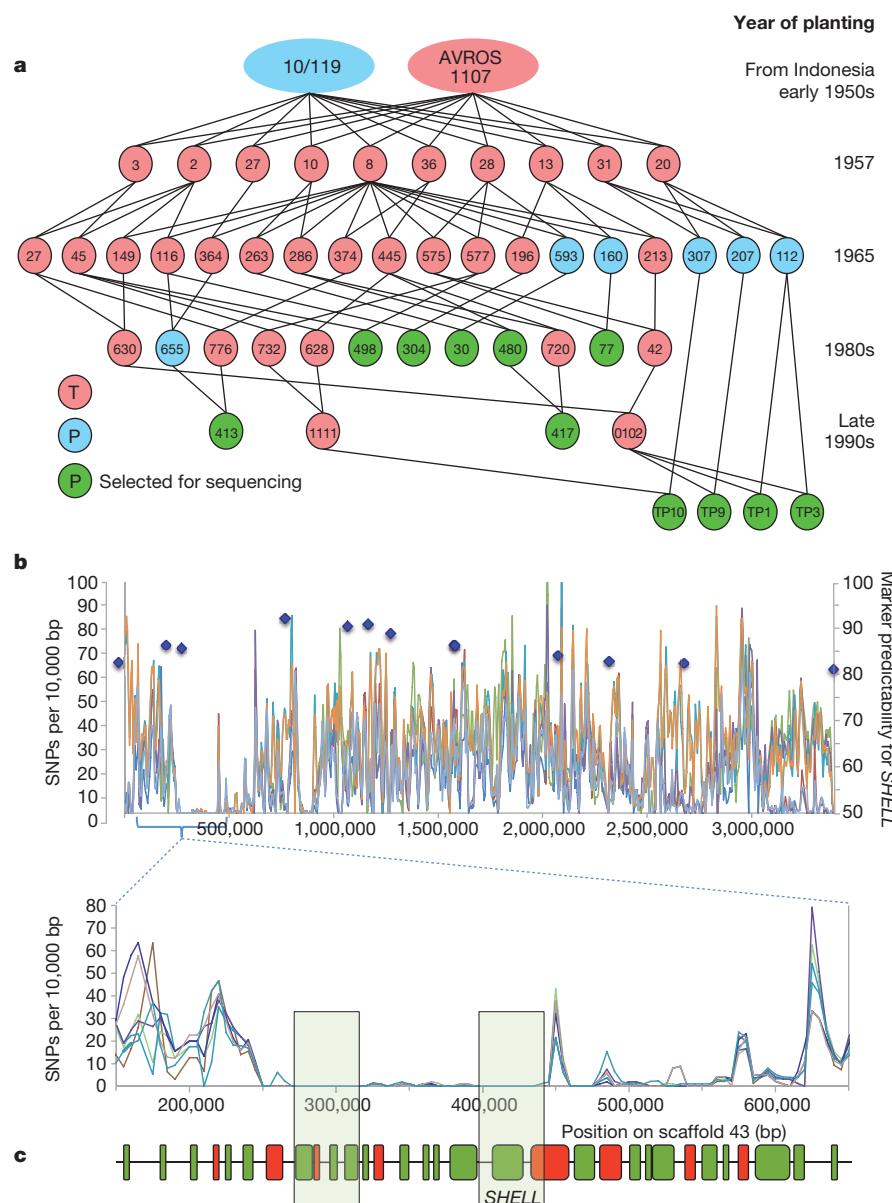


Figure 2 | Homozygosity mapping of the shell trait. **a**, AVROS pedigree palms heterozygous (*tenua*, pink) and homozygous (*pisifera*, blue and green) for *SHELL* are indicated, linked to parents and progeny planted over five decades. Individually sequenced palms are indicated (green). Three additional *pisifera* palms from this extended pedigree were also individually sequenced (not shown). The remaining *pisifera* palms were sequenced as a pool. **b**, The *SHELL* locus was mapped onto scaffold p3-sc00043 (3.4 Mb), and SNP density was plotted along the scaffold (top). Markers surrounding *SHELL* are plotted

against the percentage of *pisifera* palms with the SNP haplotype in the Nigerian T128 F₁ mapping population. *SHELL* was mapped between flanking markers, denoted by blue diamonds, at 400 kb and 1 Mb. A local diversity minimum of 200 kb was found within this interval (inset). **c**, Genes (green) and transposons (red) are indicated as boxes. The local diversity minimum contained four homozygous genes (boxed), only one of which (*SHELL*) was located between the flanking markers in **b**.

Out groups:		1		58
<i>A. thaliana</i>	(STK/AGL11)	GRGKIEIKRIENSTNRQVTFCRKRNGLKKAYELSVLCDAEVALIVFSTRGRLYEYAN		
<i>A. thaliana</i>	(SHP1/AGL1)T.....	VI	
<i>O. sativa</i>	(OsMADS13)	...R.....S.....	S.....S.	
Poplar	(MEF2-like)T.....	S.....S.	
Grape	(MADS5)T.....	S...V...S.	
Tomato	(TAGL1)T.....	S.....S.	
Peach	(SHP-like)T.....		
Oil palm, <i>dura</i> allele:				
<i>E. oleifera</i>	(0.211/2460)T.S.....	S.....	
<i>E. guineensis</i>	(0.212/70)T.S.....	S.....	
Nigeria	(FK7/005 - T128)T.S.....	S.....	
Nigeria	(GGN23)T.S.....	S.....	
Angola	(0.311/368)T.S.....	S.....	
Madagascar	(0.240/01)T.S.....	S.....	
Tanzania	(0.256/235)T.S.....	S.....	
Oil palm, <i>pisifera</i> allele:				
Nigeria	(UP0323 - T128)T.S.....P.....	S.....	
Nigeria	(GGN37)T.S.....P.....	S.....	
Angola	(0.311/460)T.S.....P.....	S.....	
Angola	(0.311/191)T.S.....N.....	S.....	
Congo	(0.182/77)T.S.....N.....	S.....	
Congo	(AVROS)T.S.....N.....	S.....	
Tanzania	(0.256/238)T.S.....N.....	S.....	

Figure 3 | Sequence diversity of *SHELL*. Multiple sequence alignment of *Arabidopsis* STK (AEE8281.1) and SHP1 (P29381.1) proteins, rice protein OsMADS13 (AAF13594.1), poplar predicted protein (XP_002327282.1), grape MADS-box protein 5 (XP_002281890.1), tomato TAGL1 (AAM33101.2) and peach SHATTERPROOF-like (ABG75908.1) (out groups); *Elaeis oleifera*, *Elaeis guineensis* and exemplars of Deli *dura* (oil palm, *dura* allele); and *pisifera*

nuclear localization, and direct DNA binding^{18,19}, whereas the substitution by a proline only two amino acid residues amino-terminal to this position would disrupt the α -helix that is involved in MADS dimerization and DNA binding¹⁹. Analysis of an additional 336 palms was used to validate these alleles within established phenotyping norms (Methods). These included four *pisifera* palms from introgression trials of *sh*^{MPOB} into *tenera* carrying the *sh*^{AVROS} allele. Sequencing confirmed that these four palms were heteroallelic, indicating that the two alleles failed to complement, and confirming the identity of the gene (Methods). To explore segregation in *E. guineensis* populations further, the *SHELL* exon 1 sequence was generated from a diversity panel of 379 palms representing nine distinct wild oil palm populations collected from Angola⁴, Madagascar⁴, Nigeria⁴ and Tanzania⁴, a subset of a 110,000-accession seed bank collected over the past five decades⁴ (Methods). We found that all palms evaluated carried either the *Sh*^{DeliDura}, *sh*^{AVROS} or *sh*^{MPOB} alleles in exon 1 (Fig. 3 and Supplementary Fig. 3).

Analysis of 10,916,126 RNA-seq reads from 22 different libraries¹² revealed only 159 reads matching *SHELL*, all of which were found in just four libraries: from whole florets one day after anthesis, kernels 10 and 15 weeks, and mesocarp 15 weeks after anthesis (WAA). *In situ* hybridization was performed on fruits between 1 and 5 WAA (Fig. 4), at the earliest stages of shell formation^{3,20}. Uniform but weak hybridization signals were detected in the mesocarp of both the thick-shelled (*dura*) and shell-less (*pisifera*) fruit forms (Fig. 4a, c), but very strong signals were detected in the outer layers of the developing kernel only in the *dura* form (Fig. 4a), consistent with the function of *SHELL*. The oil palm shell is heavily lignified (Fig. 1b), forming the endocarp, and surrounds the kernel (or pit). It is not found in *pisifera* palms, which are often female sterile.

The function and expression of the *SHELL* gene is conserved in higher plants¹⁶. In *Arabidopsis thaliana*, SHATTERPROOF (SHP1) and STK are type II MADS-box proteins of the C and D class, respectively, and form a network of transcription factors that control differentiation of the ovule, seed and lignified endocarp²¹. In tomato, SHP1 homologues

from Nigerian T128 and Congo AVROS (oil palm, *pisifera* allele). The *pisifera* fruit form is caused by two disruptive SNPs that affect a highly conserved amino acid motif in the MADS-box DNA binding and dimerization domain. Gene names (out group examples) or sample identifiers (oil palm examples) are in parentheses.

control fleshy fruit expansion in the endocarp²², and in peaches, which are also drupes, homologues of both genes have been implicated in lignified split-pit formation^{23,24}. In rice, the *SHELL* orthologue is OsMADS13 (Supplementary Fig. 4), a homologue of *Arabidopsis thaliana* STK²⁵ and SHP1 (ref. 26) that controls ovule identity, so that mutants are female sterile²⁷. STK and SHP1 bind to DNA as heteromultimers

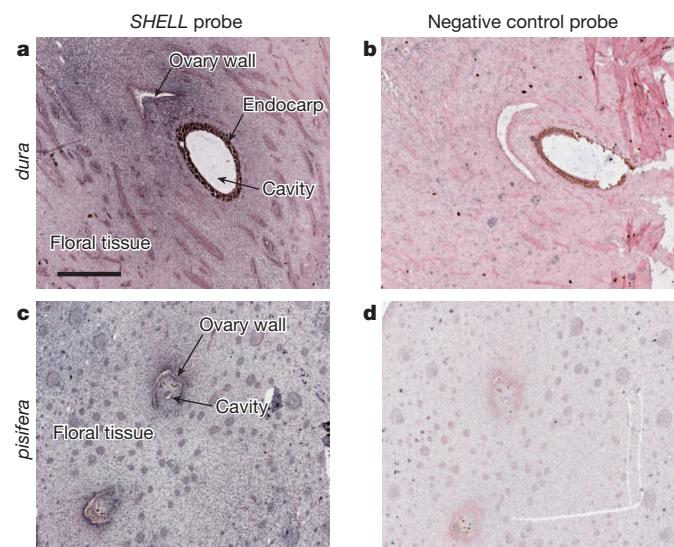


Figure 4 | Expression of *SHELL* in early oil palm fruit. a–d, Longitudinal sections of 1 WAA *dura* fruit (a, b) and cross sections of 5 WAA *pisifera* fruit (c, d) were hybridized to a pool of two unique *SHELL* locked nucleic acid (LNA) probes (a, c) or to a scrambled sequence LNA negative control probe (b, d). Colourimetric *in situ* hybridization probe detection is blue, and all sections were counterstained red. Structural hallmarks of fruits are indicated by labelled arrows in a and c. Scale bar, 0.8 mm.

Table 1 | SHELL mutations and protein interactions

Activation domain fusion	Binding domain fusion	Interaction
<i>Sh</i> ^{DeliDura}	<i>Sh</i> ^{DeliDura}	-
<i>sh</i> ^{MPOB}	<i>Sh</i> ^{DeliDura}	-
<i>sh</i> ^{AVROS}	<i>Sh</i> ^{DeliDura}	-
<i>OsMADS24</i>	<i>OsMADS24</i>	+
<i>OsMADS24</i>	<i>Sh</i> ^{DeliDura}	+
<i>OsMADS24</i>	<i>sh</i> ^{MPOB}	-
<i>OsMADS24</i>	<i>sh</i> ^{AVROS}	+

Pisifera fruit form mutations disrupt the interaction of SHELL and OsMADS24. '+' denotes dimerization detected in yeast two-hybrid assay; '-' denotes no interaction detected. Experiments were performed as described (Methods). Complete data sets are provided in Supplementary Fig. 5 and Supplementary Table 1.

with SEPALLATA (SEP) MADS-box proteins, including OsMADS24 in rice^{16,28}, and the highly conserved MADS domain is involved in both DNA binding and dimerization^{18,19}. We postulated that the mutations we detected in SHELL could account for the remarkable single-gene heterosis exhibited in *tenera* palms⁷ if they disrupted heterodimerization, as well as DNA binding^{16,18}. In yeast two-hybrid assays, we found that the *Sh*^{DeliDura} and *sh*^{AVROS} allelic forms of SHELL, but not the *sh*^{MPOB} form, interacted as a heterodimer with OsMADS24. Furthermore, OsMADS24 homodimerized with itself, but SHELL forms did not homodimerize with each other (Table 1, Supplementary Fig. 5 and Supplementary Table 1). If productive heterodimers compete with non-productive dimers in heterozygotes, this would neatly account for hybrid vigour according to theoretical models²⁹. Overdominance at a single locus accounts for similarly remarkable increases in hybrid yield in tomato³⁰.

The unravelling of the genetic basis for the shell-less phenotype paves the way for designing molecular strategies for genotyping trees that breed true for the phenotype, and modulating SHELL activity for desired fruit forms. A marker for SHELL could be used by seed producers to reduce or eliminate *dura* contamination, and to distinguish the *dura*, *tenera* and *pisifera* plants in the nursery long before they are field planted—the advantage here being that they could be planted separately based on the shell trait. This is useful as the *pisifera* palms have vigorous vegetative growth, and planting them in high density encourages male inflorescence development and pollen production. As foreseen by the breeder A. Devuyst in the Belgian Congo (in the same year that the structure of DNA was solved), accurate genotyping for enhanced oil yields will optimize and ultimately reduce the acreage devoted to oil palm plantations², providing an opportunity for conservation and restoration of dwindling rainforest reserves⁶.

METHODS SUMMARY

Mapping population. The mapping population was derived from the controlled self-pollination of the high iodine value (a measure of the level of unsaturation) *tenera* palm (T128) from the Malaysian Palm Oil Board (MPOB) Nigerian germplasm collection. Ripe fruit bunches were used for phenotyping the three fruit forms. DNA from young leaves was purified, and RFLP and SSR analyses were performed as described. SNPs were interrogated using the Infinium iSelect II assay and GenomeStudio v2010.3 (Illumina). Map construction was carried out with JoinMap v4. SNP markers from the genetic map were compared to the draft reference *pisifera* genome sequence, and highest probability SNPs were located on scaffold 43 of build 3 (p3-sc00043). Additional scaffold 43 SNPs were designed and evaluated in the mapping population using the MassArray iPlex assay (Sequenom).

Homozygosity mapping. Individual palms from the AVROS pedigree were obtained from MPOB, Kulim and Sime Darby plantations. The genomes of 14 AVROS *pisifera* palms were independently sequenced, and 29 additional AVROS *pisifera* palms were sequenced as a pool using the HiSeq 2000 instrument (Illumina). Individual reads from each tree and the pool were mapped to scaffold 43 from the reference genome assembly, and SNPs were identified and summed over 10-kb intervals. This scaffold was annotated for genes, and those with the highest homozygosity were screened for amino acid changes between *pisifera* and *dura*.

SHELL sequencing. PCR amplicons of SHELL exon 1 were sequenced twice each in the forward and reverse directions, using the ABI3730 instrument (Life Technologies) (Methods).

SHELL *in situ* hybridization. Longitudinal and cross sections were hybridized with locked nucleic acid digoxigenin-labelled oligonucleotide probes. Slides were developed with alkaline phosphatase and counterstained with Nuclear Fast Red (Methods).

Yeast two-hybrid assays. Oil palm *Sh*^{DeliDura}, *sh*^{MPOB}, *sh*^{AVROS} and rice OsMADS24 sequences were each synthesized as two overlapping gBlock gene fragments (Integrated DNA Technologies), combined using Gibson Assembly Master Mix (NEB), and ligated in-frame into MatchMaker Gold Yeast Two-Hybrid System vectors (Clontech). Constructs included the coding sequences of the MADS-box, I and K domains of each protein (Methods). Clones were sequence verified by Sanger sequencing (ABI). Yeast (AH109) co-transformations and interaction assays were performed following MatchMaker Gold protocols (Clontech).

Full Methods and any associated references are available in the online version of the paper.

Received 23 November 2012; accepted 7 June 2013.

Published online 24 July 2013.

- Janssens, P. Le palmier à huile au Congo Portugais et dans l'enclave de Cabinda. Descriptions de principales Variétés de Palmier (*Elaeis guineensis*). *Bull. Agric. Congo Belge* **18**, 29–92 (1927).
- Devuyst, A. Selection of the oil palm (*Elaeis guineensis*) in Africa. *Nature* **172**, 685–686 (1953).
- Hartley, C. *The Oil Palm* 47–94 (Longman, 1988).
- Rajanaidu, N. *et al.* in *Advances in Oil Palm Research* (eds Basiron, Y., Jalani, B. S. & Chan, K. W.) 171–237 (Malaysian Palm Oil Board, 2000).
- Corley, R. H. V. & Tinker, P. B. *The Oil Palm* 4th edn, 1–26 (Blackwell Science, 2003).
- Danielsen, F. *et al.* Biofuel plantations on forested lands: double jeopardy for biodiversity and climate. *Conserv. Biol.* **23**, 348–358 (2009).
- Beirmaert, A. & Vanderwegen, R. Contribution à l'étude génétique et biométrique des variétés d'*Elaeis guineensis* Jacq. *Publ. Inst. Nat. Etude Agron. Congo Belge. Ser. Sci.* **27**, 1–101 (1941).
- Goddin, R. Observation de la production de palmiers sélectionnés à Mongana (Equateur). *Bull. Agric. Congo Belge* **21**, 1263 (1930).
- Billotte, N. *et al.* Microsatellite-based high density linkage map in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* **110**, 754–765 (2005).
- Mayes, S., Jack, P. L., Corley, R. H. & Marshall, D. F. Construction of a RFLP genetic linkage map for oil palm (*Elaeis guineensis* Jacq.). *Genome* **40**, 116–122 (1997).
- Moretzsohn, M. C., Nunes, C. D. M., Ferreira, M. E. & Grattapaglia, D. RAPD linkage mapping of the shell thickness locus in oil palm (*Elaeis guineensis* Jacq.). *Theor. Appl. Genet.* **100**, 63–70 (2000).
- Singh, R. *et al.* Oil palm genome sequence reveals divergence of interfertile species in Old and New Worlds. *Nature* <http://dx.doi.org/10.1038/nature12309> (this issue).
- Rajanaidu, N., Rao, V., Abdul Halim, H. & Ong, A. S. H. Genetic resources. New developments in oil palm breeding. *Elaeis* **1**, 1–10 (1989).
- Gschwend, M. *et al.* A locus for Fanconi anaemia on 16q determined by homozygosity mapping. *Am. J. Hum. Genet.* **59**, 377–384 (1996).
- Lander, E. S. & Botstein, D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* **236**, 1567–1570 (1987).
- Favarro, R. *et al.* MADS-box protein complexes control carpel and ovule development in *Arabidopsis*. *Plant Cell* **15**, 2603–2611 (2003).
- Pinyopich, A. *et al.* Assessing the redundancy of MADS-box genes during carpel and ovule development. *Nature* **424**, 85–88 (2003).
- Huang, H. *et al.* DNA binding properties of two *Arabidopsis* MADS domain proteins: binding consensus and dimer formation. *Plant Cell* **8**, 81–94 (1996).
- Imrnink, R. G., Kaufmann, K. & Angenent, G. C. The 'ABC' of MADS domain protein behaviour and interactions. *Semin. Cell Dev. Biol.* **21**, 87–93 (2010).
- Bhasker, S. & Mohankumar, C. Association of lignifying enzymes in shell synthesis of oil palm fruit (*Elaeis guineensis*-dura variety). *Indian J. Exp. Biol.* **39**, 160–164 (2001).
- Dinneny, J. R. & Yanofsky, M. F. Drawing lines and borders: how the dehiscent fruit of *Arabidopsis* is patterned. *Bioessays* **27**, 42–49 (2005).
- Vrebolj, J. *et al.* Fleshy fruit expansion and ripening are regulated by the Tomato SHATTERPROOF gene TAGL1. *Plant Cell* **21**, 3041–3062 (2009).
- Tani, E., Polidoros, A. N. & Tsafaris, A. S. Characterization and expression analysis of FRUITFULL- and SHATTERPROOF-like genes from peach (*Prunus persica*) and their role in split-pit formation. *Tree Physiol.* **27**, 649–659 (2007).
- Tani, E. *et al.* Characterization and expression analysis of AGAMOUS-like, SEEDSTICK-like, and SEPALLATA-like MADS-box genes in peach (*Prunus persica*) fruit. *Plant Physiol. Biochem.* **47**, 690–700 (2009).
- Kramer, E. M., Jaramillo, M. A. & Di Stilio, V. S. Patterns of gene duplication and functional evolution during the diversification of the AGAMOUS subfamily of MADS box genes in angiosperms. *Genetics* **166**, 1011–1023 (2004).
- Zahn, L. M. *et al.* Conservation and divergence in the AGAMOUS subfamily of MADS-box genes: evidence of independent sub- and neofunctionalization events. *Evol. Dev.* **8**, 30–45 (2006).
- Dreni, L. *et al.* The D-lineage MADS-box gene *Osmads13* controls ovule identity in rice. *Plant J.* **52**, 690–699 (2007).
- Cui, R. *et al.* Functional conservation and diversification of class E floral homeotic genes in rice (*Oryza sativa*). *Plant J.* **61**, 767–781 (2010).
- Veitia, R. A. & Vaiman, D. Exploring the mechanistic bases of heterosis from the perspective of macromolecular complexes. *FASEB J.* **25**, 476–482 (2011).

30. Krieger, U., Lippman, Z. B. & Zamir, D. The flowering gene *SINGLE FLOWER TRUSS* drives heterosis for yield in tomato. *Nature Genet.* **42**, 459–463 (2010).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank J. Jansen for assistance with JoinMap, and Beijing Genome Institute, The Genome Institute at Washington University, Tufts University Core Facility and the Arizona Genome Institute for mapping and genome sequencing. We thank Kulim, Sime Darby, FELDA Agricultural Services, United Plantations and Applied Agricultural Resources for providing materials and phenotype information of individual palms. Phylogeny, Inc. performed *in situ* hybridization services. Creative Biolabs performed yeast two-hybrid co-transformations and interaction assays. We thank J. Birchler and Z. Lippman for discussions on hybrid vigour. We appreciate the constant support of Y. M. Choo, and the Ministry of Plantation Industries and Commodities, Malaysia. The project was endorsed by the Cabinet Committee on the Competitiveness of the Palm Oil Industry (CCPO) and funded by the Malaysian Palm Oil Board. R.A.M. is supported by a grant from National Science Foundation 0421604 ‘Genomics of Comparative Seed Plant Evolution’.

Author Contributions R.S. initiated work on the *SHELL* marker/gene. R.S., E.-T.L.L., M.O.-A. and R.S. conceptualized the research programme. R.S., E.-T.L.L., M.O.-A., R.N.,

M.A.M., N.L., S.W.S., J.M.O., R.S. and R.A.M. developed the overall strategy, designed experiments and coordinated the project. R.A.M. conceptualized the homozygosity mapping strategy. R.S., R.N. and R.A.M. identified samples for homozygosity mapping. R.N. and M.D.A. developed and maintained the mapping population and assisted in phenotyping. R.S., M.O.-A., L.C.-L.O., N.-C.T., J.N., N.L., M.A.B., B.B., A.V.B., C.W., J.M.O. and R.S. conducted laboratory experiments and data analyses. R.S. and L.C.-L.O. constructed the genetic map. E.-T.L.L., R.R., K.-L.C., M.A.H., N.A., S.W.S., M.H., A.V.B. and C.W. performed bioinformatics analyses. N.L., A.V.B., C.W., J.M.O., R.S. and R.A.M. resequenced the candidate gene and characterized the mutations. R.S., E.-T.L.L., M.O.-A., R.N., N.L., S.W.S., J.M.O., R.S. and R.A.M. prepared and revised the manuscript.

Author Information *E. guineensis* and *E. oleifera* genome sequences have been deposited at DDBJ/EMBL/GenBank under the accessions ASJS00000000 and ASIR00000000, respectively. Gene sets are available at <http://genomsawit.mpopb.gov.my>. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details accompany the full-text HTML version of the paper at www.nature.com/nature. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.S. (raviga@mpob.gov.my) or R.A.M. (martiens@cshl.edu).

METHODS

Plant materials and germplasm collection. Oil palm germplasm materials used in this study were collected through bilateral agreements between Malaysia and the countries of origin, and were in accordance with the Convention on Biological Diversity 1992. The reference Deli *dura* palm was Malaysian Palm Oil Board (MPOB) 0.212/70, the reference *pisifera* palm was MPOB 0.182/77, and the reference *E. oleifera* palm was MPOB 0.211/2460. The mapping family used for generating the genetic linkage map was derived from controlled self-pollination of the high iodine value *virescens tenera* palm, T128 (accession number MPOB 371), from the MPOB Nigerian germplasm collection¹³. A total of 241 palms were originally planted from 1993 to 1997 at several locations in Malaysia, namely the MPOB-UKM Research Station Bangi at Selangor, MPOB Ulu Paka Research Station at Terengganu, MPOB Keratong Research Station at Pahang, MPOB Lahad Datu Research Station at Sabah, United Plantations at Perak and the FELDA Research Station at Pahang. Of the 241 palms, 240 were still available for marker analysis. The T128 palm is relevant to the breeding programme as it has outstanding attributes in terms of oil yield, fatty acid composition and low height increment. It has been crossed with at least six *dura*, eight *tenera* and three *pisifera* palms for progeny testing and widening the genetic base. A collection of 379 palms (287 *dura*, 86 *tenera* and 6 *pisifera* scored fruit forms) from Angola, Congo, Madagascar, Nigeria and Tanzania were taken from the diversity panel maintained by MPOB, and screened for new alleles as described. Unopened leaf samples (spear leaves) were collected from individual palms and immediately frozen in liquid nitrogen and stored at -80 °C until DNA preparation. DNA was extracted and purified from the leaf samples as previously described³¹.

Whole-genome sequencing and assembly. An AVROS *pisifera* palm was sequenced to high coverage¹² on the 454 GS FLX Titanium XL+ next-generation sequencing platform (454/Roche). Sequence reads were generated from DNA fragment libraries and from a complex series of linker libraries, in which read ends span fragment sizes ranging from 0.75 to >20 kb. Sequence reads were assembled with the Newbler assembler (Roche 454), producing a reference sequence of the oil palm genome¹². Scaffolds from the reference assembly containing markers genetically mapped in the *SHELL* interval were identified. A BAC physical map was constructed from a ten-fold BAC library constructed from the same AVROS *pisifera* used to generate the reference sequence. BAC end sequences were also generated from each BAC in the library using standard Sanger sequencing on the 3730 sequencing platform (Life Technologies). BAC end sequences were assembled into the reference genome with the Newbler assembler. A minimum tiling path of BAC clones spanning the *SHELL* interval was selected, and BAC clones in the tiling path were sequenced in pools to high coverage with 454 XL technology. The BAC pool sequences were assembled and merged with the original 454 whole-genome shotgun sequence data and all BAC end sequences. Improved scaffold coverage and scaffold length spanning the *SHELL* interval was produced.

Genetic mapping. A total of 241 progeny palms were derived from self-pollination of the Nigerian *tenera* palm T128 (ref. 4), of which 240 palms were available throughout the study. Two palms could not be phenotyped accurately, whereas 124 *tenera*, 46 *pisifera* and 68 *dura* palms were phenotyped with high confidence at the time of genetic map construction. All 240 progeny were scored for 4,451 SNP markers using the Infinium iSelect Assay (Illumina), as well as 3 RFLP markers³¹, 4 SSR markers³¹ and 193 additional SSR markers developed from the oil palm genome sequence¹². The linkage map was then constructed using JoinMap 4.0 (<http://www.kyazma.nl/index.php/mc.JoinMap>). The genotype data was formatted as required for mapping according to an *F*₂ population. Markers showing segregation profile of 1:2:1 were used in the map construction.

Two sets of the genotype data were then created, in which one set is the converse of the other to account for phase differences in the T128 'selfed' *F*₂ population. Markers that exhibited severe distortion ($P < 0.0001$) and markers having more than 10% missing data were excluded. Both sets of genotype data were then grouped at a recombination frequency of ≤ 0.2 . Next, 18 nodes were selected to create 18 initial groups for calculating the linkage groups. The linkages were calculated and loci ordered based on the maximum likelihood algorithm. None of the markers showed severe distortion ($P < 0.0001$). The few distorted markers observed were significant at $P < 0.05-0.1$. These markers were removed from further analysis when necessary. Markers exhibiting nearest neighbour stress value > 2 (cM) were identified and excluded from the analysis. Markers contributing to insufficient linkages were also determined and removed. The T128 co-dominant map constructed consisted of 16 groups, and *SHELL* was placed on linkage group 7, within the expected region (Supplementary Fig. 1).

The SNP markers surrounding *SHELL* were subsequently mapped by sequence similarity to a 3.4-Mb assembly scaffold 43 (p3-sc00043). A tiling path of BAC contigs corresponding to scaffold 43 was selected from a high-information content physical map of *pisifera* and sequenced¹². An additional 50 SNP assays were designed from an improved assembly (p5-sc00060) corresponding to scaffold

43. Thirty additional SNP markers were also designed from scaffolds p3-sc00191, p3-sc00203 and p3-sc02216, which were also associated with markers on linkage group 7, and reassembled as p5-sc00263. These 80 SNP markers (designated as SNPE) were genotyped in the T128 selfed population using the Sequenom MassArray iPLEX platform, and 63 were polymorphic giving a final co-dominant genetic linkage map consisting of 818 markers (719 SNP (inclusive of SNPE), 96 SSR, 2 RFLP markers and *SHELL*) in 16 linkage groups. *SHELL* remained in linkage group 7, together with the 63 SNPE markers developed from the selected scaffolds. The final size of linkage group 7 is approximately 182 cM, with an average of 1.2 cM between two adjacent markers (Supplementary Fig. 1).

Single marker mapping was used to determine the recombination breakpoints surrounding *SHELL* in each individual palm. To obtain a genetic interval containing *SHELL*, palms with multiple breakpoints in the region and/or inconsistent haplotypes were discarded in case of mis-genotyping. Markers flanking *SHELL* were selected using these conservative criteria and found to lie 450 kb apart on scaffold p3-sc00043 (Fig. 2b, c).

Fruit form phenotyping. Oil palm trees were grown to maturity in open plantations, making accurate phenotyping of fruit form for some samples difficult owing to variation in fertility, yield and environment. Several fruits were harvested from each palm, and shell thickness and fruit form was determined using established criteria³. To assess the accuracy of the fruit form phenotypic data used in the study, we reviewed 460 phenotype calls that were made between 2003 and 2012. In this period, up to three independent attempts were made to visually determine the fruit form phenotypes of 340 palms, 241 of which were from the T128 selfed population used to map *SHELL*, and 99 from a different population for which re-phenotyping data were available. In the data, ambiguous calls were made 26 times (or 5.7% of total phenotype determinations) where breeders were unsure of the fruit form phenotype.

Homozygosity mapping. A total of 43 individual *pisifera* palms from the AVROS pedigree (originating from Congo) were sourced from the MPOB, Sime Darby and Kulim plantations. Fourteen of the AVROS *pisifera* palms were independently sequenced, whereas the DNA from the remaining 29 palms was pooled for sequencing (pool 1). Whole-genome shotgun sequence data was generated on the HiSeq 2000 (Illumina). Individual trees and pools of trees were sequenced to 20× and 40× raw sequence coverage, respectively. Individual reads were mapped to sequence scaffolds from the *pisifera* reference genome assembly, and highest probability SNPs were located on scaffold 43 of the *pisifera* build 3 (p3-sc00043). In each of the 14 genome sequences, SNPs were summed over 10-kb windows along the scaffold and plotted against map location. This scaffold was computationally annotated for genes by comparing to public databases for the *A. thaliana* and rice genomes using the NCBI BLAST similarity searching tool. The genes with the highest homozygosity within the predicted interval from the genetic map were screened for putative function and for nucleic and amino acid changes between the *pisifera* and *dura* lines, as well as for amino acid differences in other species.

Validation and genotyping. DNA sequencing of the eight exons of the *SHELL* gene was carried out for all palms of the mapping family as well as the AVROS *pisifera* palms used in homozygosity mapping. One of the 14 sequenced *pisifera* palms (TP10) proved to be heterozygous for the *sh*^{AVROS} and *sh*^{MPOB} haplotypes, due to contamination in a breeding trial, and was not used for the homozygosity analysis. PCR primers were designed based on the reference *pisifera* genome sequence to amplify the entirety of *SHELL* exon 1. *SHELL*-specific primer sequences were 5'-TTGCTTTAATTGCTGAATACC-3' (forward primer upstream of exon 1) and 5'-TTGGATCAGGGATAAAAGGAAGC-3' (reverse primer downstream of exon 1). Primer sequences were confirmed to be unique in the reference *pisifera* genome, and to avoid any identified polymorphic nucleotides. A 5' M13 forward sequence tag (5'-GTTTCCCAGTCACGACGTGTA-3') was added to the exon 1 forward PCR primer. A 5' M13 reverse sequence tag (5'-AGGAAACAGCTATGACCAT-3') was added to the exon 1 reverse primer. *SHELL* exon 1 was amplified from genomic DNA and PCR amplification was performed using 20 ng of purified genomic DNA under standard PCR amplification conditions. Amplicons were treated with exonuclease I and shrimp alkaline phosphatase to remove unincorporated primers and deoxynucleotides. An aliquot of each amplicon was sequenced using a M13 forward primer on an ABI3730 instrument under standard conditions. Each amplicon was sequenced twice in the forward direction. An aliquot of each amplicon was additionally sequenced using the M13 reverse primer. Each amplicon was sequenced twice in the reverse direction. All sequencing data were aligned to the reference *pisifera* genome sequence. Data were analysed to determine the genotype at each of the two SNP positions identified to be associated with the *pisifera* fruit form.

We sequenced *SHELL* exon 1 from 336 individual palms from the T128 mapping population, the samples used to construct homozygosity maps, and a collection of palms in crosses with advanced lines (100 *pisifera*, 148 *tenera*, 86 *dura* and 2 with ambiguous phenotype). A total of 323 (96.7%) had *SHELL* genotypes concordant

with their phenotype, and 11 (3.3%) had discordant phenotypes, reflecting the accuracy of phenotyping in the plantation (see above). *SHELL* exon 1 was also sequenced from all four *pisifera* palms derived from T × T crosses between *sh*^{MPOB} and *sh*^{AVROS} (FELDA AA and MPOB PK575), and proved to be heteroallelic as predicted. A further three *pisifera* palms were also found to be heteroallelic, including TP10 which was sequenced completely and proved to be a contaminant in the AVROS pedigree (see above). The other two heteroallelic palms were probably similarly contaminated.

A second attempt was made to phenotype and resequence the 11 apparently discordant trees, enabling the re-evaluation of nine trees for fruit form. The second phenotype call of one palm was concordant with the genotype prediction, whereas the second phenotype calls of seven palms were ambiguous, and one palm retained the original phenotype. This palm was re-genotyped and proved to have a consistent genotype. It is plausible that the nine palms remaining (or 2.7% of the genotyped population) had been misphenotyped originally, given that fruit form phenotyping error is believed to be in excess of 5% (see above), highlighting the need for a molecular assay that more accurately predicts fruit form. This assumption was confirmed in six of the nine palms for which haplotypes were available, as haplotypes were consistent with genotype not phenotype, ruling out recombination as an explanation for discrepancy. The map expansion immediately around the *SHELL* gene is similarly explained by misphenotyped palms. That is, four SNP markers flanking the *SHELL* gene, have 9 recombinants (of the 238 palms), which include 6 or the 9 misphenotyped palms from sequencing of the *SHELL* exon 1. **In situ hybridization.** Tissue sections (5 µm) were mounted on Superfrost slides, dried overnight, and baked at 60 °C. Sections were deparaffinized in xylene, immersed in 100% ethanol and dried. The sections received light treatment with protease. Two LNA probes to the target messenger RNA were designed and supplied by Exiqon (5'-DIG-ATTAACAAGCAGCGACATACTT-3' and 5'-DIG-TTGATGGTGTGAATAGTGTGT-3'). A Scramble-miR negative control LNA probe was also provided by Exiqon (5'-DIG-GTGT AACACGTCTATACGC CCA-3'). Optimized probe cocktail solution in Exiqon hybridization buffer was placed on the tissue sections. The sections were covered with polypropylene coverslips and heated to 60 °C for 5 min followed by hybridization at 37 °C overnight. Sections were washed in high stringency solution (0.2× SSC with 2% BSA) at 60 °C for 10 min. The LNA probes were detected using alkaline phosphatase conjugate (NBT/BCIP, blue precipitate). Sections were counterstained with Nuclear Fast

Red. Sections were rinsed and mounted with coverslips. Aperio scans of the slides were made and images were extracted with ImageScope software.

Yeast two-hybrid assays. The coding sequences for oil palm *Sh*^{DeliDura}, *sh*^{MPOB}, *sh*^{AVROS} and rice *OsMADS24* were synthesized as two approximately 300-bp gBlocks each that overlapped by 30 bp (Integrated DNA Technologies). Gibson assembly of the two fragments was performed using kit manufacturer's protocols (NEB). EcoRI and BamHI sites were added to the gBlock sequences for simple ligation into MatchMaker Gold Yeast Two-Hybrid vectors. Each sequence was cloned into both the binding domain vector, pGBT7, and the activation domain vector, pGAD7. *SHELL* sequences encoded amino acids 2–175, including the entire MADS-box, I and K domains. The C domain was excluded from yeast two-hybrid constructs to avoid auto-activation of selection genes in the yeast two-hybrid system. The *Sh*^{DeliDura} peptide sequence encoded by the vectors was: GRGKIEIKRIENTTSRQVTFCRKRRNGLLKKAYEVSVDCAEVALIVFSSRGRL YEYANNSIRSTIDRYKKACANSSNSGATIE~~I~~NSQQYYQQESAKLRHQIQLQN ANRHLMGEALSTLTVKELQLENRLERGIRTRISKKHELLFAEIEYMQKREV ELQNDNMYLRAKIAEN. The *sh*^{MPOB} peptide sequence encoded by the vectors was identical to the above sequence, with the exception that the underlined Leu residue (L) was converted to Pro (P). The *sh*^{AVROS} peptide sequence encoded by the vectors was identical to the above sequence, with the exception that the underlined Lys residue (K) was converted to Asn (N). *OsMADS24* sequences encoded amino acids 2–177, including the entire MADS-box, I and K domains, but excluding the C domain. The *OsMADS24* sequence encoded by the vectors was: GRGRVELKRIENKINRQVTFAKRRNGLLKKAYEVSVDCAEVALIIFSNRGK LYEFCSGQSMTRTLERYQKF SYGGPD~~T~~AIQNKENELVQSSRNEYLKLKARV ENLQRTQRNLLGEDLGLGIKELEQLEKQLDSSLRHIRSTRQHMLDQLTD LQRREQMLCEANKCLRRKLEES. Auto-activation control tests were performed by transforming each fusion vector into yeast alone, and each vector showed no auto-activation of selection reporter genes. Co-transformations were performed for all 16 pairwise combinations of BD and AD vectors and scored for growth on SD-Leu-Trp, SD-Leu-Trp-His, SD-Leu-Trp-His-Ade and X-gal media plates. Positive interactions were scored as blue co-transformants (on X-gal plate) that were able to grow on SD-Leu-Trp-His-Ade selection plates (Supplementary Fig. 5).

31. Singh, R. et al. Mapping quantitative trait loci (QTLs) for fatty acid composition in an interspecific cross of oil palm. *BMC Plant Biol.* **9**, 114 (2009).

Maternal imprinting at the *H19–Igf2* locus maintains adult haematopoietic stem cell quiescence

Aparna Venkatraman^{1,2}, Xi C. He¹, Joanne L. Thorvaldsen³, Ryohichi Sugimura¹, John M. Perry¹, Fang Tao¹, Meng Zhao¹, Matthew K. Christenson¹, Rebeca Sanchez^{1,4}, Jaclyn Y. Yu¹, Lai Peng⁵, Jeffrey S. Haug¹, Ariel Paulson¹, Hua Li¹, Xiao-bo Zhong⁵, Thomas L. Clemens⁶, Marisa S. Bartolomei³ & Linheng Li^{1,7}

The epigenetic regulation of imprinted genes by monoallelic DNA methylation of either maternal or paternal alleles is critical for embryonic growth and development¹. Imprinted genes were recently shown to be expressed in mammalian adult stem cells to support self-renewal of neural and lung stem cells^{2–4}; however, a role for imprinting per se in adult stem cells remains elusive. Here we show upregulation of growth-restricting imprinted genes, including in the *H19–Igf2* locus⁵, in long-term haematopoietic stem cells and their downregulation upon haematopoietic stem cell activation and proliferation. A differentially methylated region upstream of *H19* (H19-DMR), serving as the imprinting control region, determines the reciprocal expression of *H19* from the maternal allele and *Igf2* from the paternal allele¹. In addition, *H19* serves as a source of miR-675, which restricts *Igf1r* expression⁶. We demonstrate that conditional deletion of the maternal but not the paternal H19-DMR reduces adult haematopoietic stem cell quiescence, a state required for long-term maintenance of haematopoietic stem cells, and compromises haematopoietic stem cell function. Maternal-specific H19-DMR deletion results in activation of the *Igf2–Igf1r* pathway, as shown by the translocation of phosphorylated FoxO3 (an inactive form) from nucleus to cytoplasm and the release of FoxO3-mediated cell cycle arrest, thus leading to increased activation, proliferation and eventual exhaustion of haematopoietic stem cells. Mechanistically, maternal-specific H19-DMR deletion leads to *Igf2* upregulation and increased translation of *Igf1r*, which is normally suppressed by *H19*-derived miR-675. Similarly, genetic inactivation of *Igf1r* partly rescues the H19-DMR deletion phenotype. Our work establishes a new role for this unique form of epigenetic control at the *H19–Igf2* locus in maintaining adult stem cells.

Our earlier studies showed that imprinted genes, including those within the *H19–Igf2* locus (Fig. 1a), are differentially expressed in haematopoietic stem and progenitor cells (HSPCs)⁷. To explore this further, we systematically analysed imprinted gene expression in quiescent-enriched long-term haematopoietic stem cells (LT-HSCs), more active short-term (ST)-HSCs and multipotent progenitor (MPP) populations (Fig. 1b)⁸. Out of 88 imprinted genes, 23 were differentially expressed in these populations. Of these 23, 15 were preferentially expressed in LT-HSCs, whereas the others were predominantly expressed in ST-HSCs and MPPs (Fig. 1c). Intriguingly, 80% of the imprinted genes with predominant expression in LT-HSCs were associated with growth restriction, including *H19*, *Cdkn1c/p57*, *Ndn*, *Rb*, *Gtl2* and *Grb10* (ref. 9). In contrast, imprinted genes expressed preferentially in ST-HSCs and MPPs, including *Ascl2*, *Peg12*, *Sfmbt2*, *Pon3*, *Atp10a* and *Osbpl5*, were associated with growth promotion and increased metabolism¹⁰ (see quantitative real-time PCR assay in Supplementary Fig. 1a).

Given the critical role of *H19* during embryonic development and its preferential expression in LT-HSCs, we considered that it plays a role

in restricting LT-HSC activation. To test this idea, we conditionally deleted H19-DMR (an epigenetic regulator that controls expression of *H19*) by breeding *H19*^{flDMR/flDMR} mice with *Mx1-Cre* mice to generate maternal (*mH19*^{ΔDMR/+}) and paternal (*pH19*^{ΔDMR/+}) allele-specific mutants (Supplementary Fig. 1b). The DMR region was deleted with 100% efficiency in LT-HSC (Supplementary Fig. 1c, e–g)¹¹. As early as 6 weeks, flow cytometric analysis showed a substantial decrease in frequency and absolute number of LT-HSCs in *mH19*^{ΔDMR/+} mice (Fig. 1d–f and Supplementary Fig. 1d), but not in *pH19*^{ΔDMR/+} mice (Fig. 1e and Supplementary Fig. 2a, b). Concurrently, we observed a significant increase in frequency and absolute number of ST-HSCs (Fig. 1d–f); however, the total number of bone marrow cells remained unchanged (Fig. 1g). By 6 months, both LT- and ST-HSCs were significantly decreased in frequency and absolute number, whereas bone marrow cellularity increased only in *mH19*^{ΔDMR/+} mice (Fig. 1d–g and Supplementary Fig. 2c, d).

Cell cycle analysis of LT-HSCs at 6 weeks after pIpC induction showed a decrease in the G0 phase fraction and a concomitant increase in the G1 phase fraction in *mH19*^{ΔDMR/+} relative to control (Fig. 1h–j). We then tested the response of *mH19*^{ΔDMR/+} mice to bone marrow damage by administering 5-fluorouracil (5FU), which eliminates active HSPCs while sparing quiescent HSCs. Surviving quiescent HSCs later replenish lost HSPCs¹² (Fig. 1k). In this context, a significant reduction in quiescent HSCs after three cycles of 5FU treatment led to deficient bone marrow recovery in the *mH19*^{ΔDMR/+} mutant compared with control (Fig. 1l, m and Supplementary Fig. 2f). Altogether, maternal but not paternal deletion of H19-DMR resulted in loss of HSC quiescence, leading to progressive loss of LT-HSCs and then ST-HSCs, accompanied by increasing progenitor cell proliferation and differentiation, thus ultimately increasing total bone marrow cellularity (Fig. 1g and Supplementary Figs 2e and 3a–d).

To characterize the phenotype functionally, we transplanted equal numbers of sorted LT-HSCs from mutants and their control littermates. We observed a significant reduction in reconstitution ability for LT-HSCs derived from *mH19*^{ΔDMR/+} but not *pH19*^{ΔDMR/+} mutants compared with controls. Although overall engraftment was reduced in primary and secondary recipients, no mature lineage bias was apparent (Fig. 2a–f). Limiting dilution analysis to quantify functional HSCs showed a 2.5-fold decrease in *mH19*^{ΔDMR/+} mutant HSCs relative to control (Fig. 2d). Reciprocal transplantation of wild-type donor cells into either *mH19*^{ΔDMR/+} or control recipients did not result in alterations in haematopoiesis (Fig. 2g, h), indicating that an intrinsic change in the *mH19*^{ΔDMR/+} mutant HSCs was the primary cause of the phenotype.

Next, we investigated whether H19-DMR controls the imprinted expression of *H19* and *Igf2* from the maternal and paternal alleles, respectively, in adult HSCs, as is observed in embryos¹¹. Our RNA-seq analysis showed differential expression of *H19* as well as *Igf2* in

¹Stowers Institute for Medical Research, Kansas City, Missouri 64110, USA. ²Centre for Stem Cell Research, Christian Medical College, Vellore 632002, India. ³Department of Cell & Developmental Biology, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA. ⁴Cell Therapy and Genetics Laboratory, Institut de Recerca Vall d'Hebron, Barcelona 08035, Spain. ⁵Department of Pharmacology, Toxicology & Therapeutics, University of Kansas Medical Center, Kansas City, Kansas 66160, USA. ⁶Center for Musculoskeletal Research, Johns Hopkins Medicine, Baltimore, Maryland 21287, USA.

⁷Department of Pathology & Laboratory Medicine, University of Kansas Medical Center, Kansas City, Kansas 66160, USA.

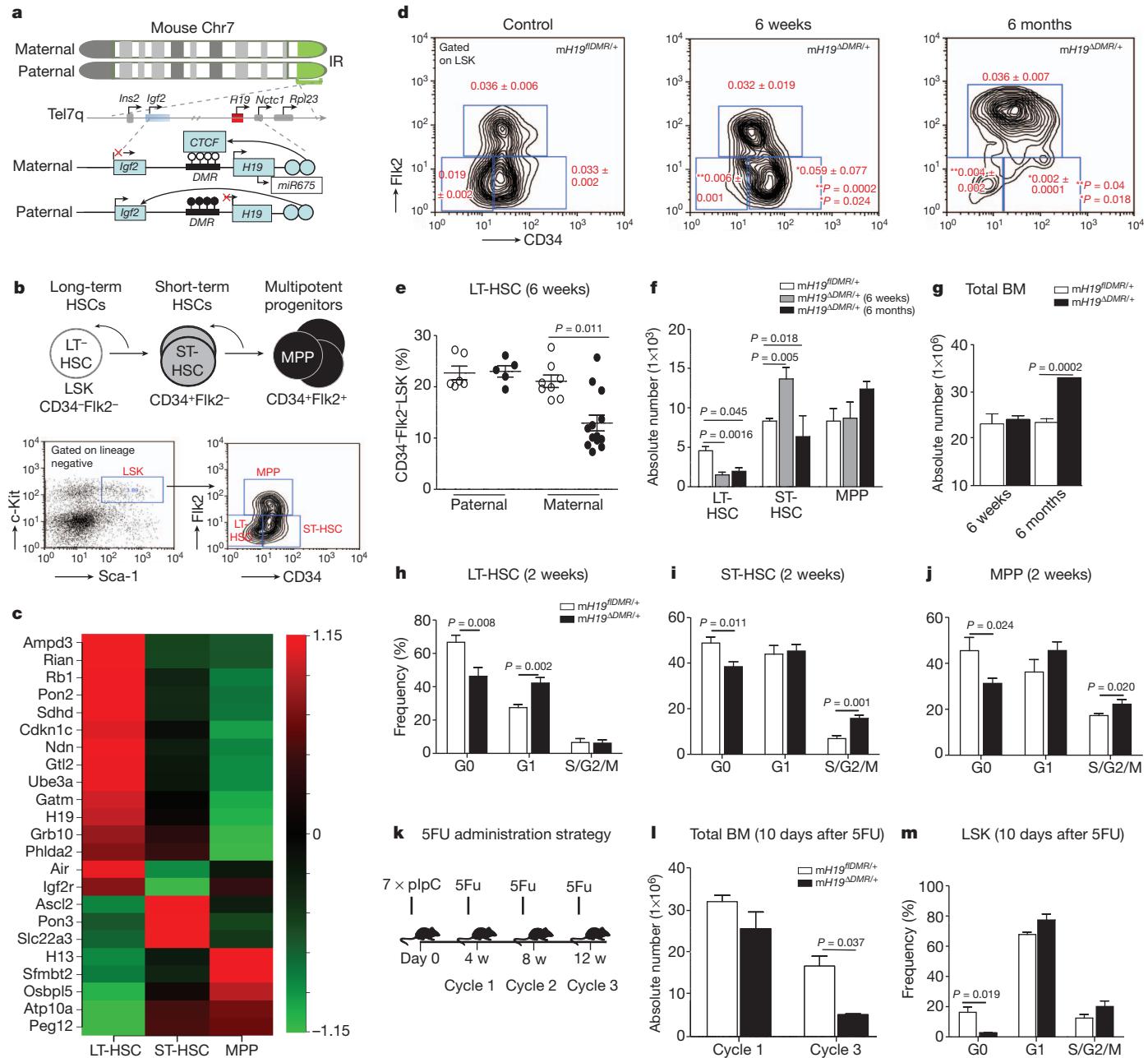


Figure 1 | Defective LT-HSCs in *mH19*^{ΔDMR/+} mice. **a**, *H19*-*Igf2* cluster. Top: red box, maternally expressed; blue box, paternally expressed; grey boxes, genes in the cluster. Bottom: unmethylated (open circles) and methylated (filled circles) CpG dinucleotides; red cross, no expression; arrow, active transcription. IR, imprinting region. **b**, Upper cartoon, hierarchical organization. Curved arrow, self-renewal. Lower image: fluorescence-activated cell sorting (FACS) plot. Left panel: gated on LSK; right panel: on CD34 and Flk2. **c**, Heat map of

imprinted genes ($n = 3$). **d**, FACS plot of *mH19*^{ΔDMR/+} ($n = 9$) and *mH19*^{fl/fl DMR/+} ($n = 8$). Asterisks and double asterisks, significant compared with control. **e**, Percentage of LT-HSCs of LSK. Absolute numbers of **f**, HSCs and **g**, total bone marrow cells. **h–j**, Cell cycle analysis 2 weeks after pIpC induction ($n = 5$). **k**, 5FU cycles. **l**, Total bone marrow cells ($n = 3$). **m**, Cell cycle analysis. Error bars, s.e.m.

HSCs (Fig. 3a, b). By crossing *H19*^{fl/fl} females with *Mus castaneus* (Cast) males, which enables parental allele-discrimination by single nucleotide polymorphism analysis, we further detected exclusive expression of *Igf2* from the paternal allele in HSCs (Fig. 3c). However, after deletion of the maternal H19-DMR, we detected *H19* downregulation and *Igf2* upregulation, which resulted from biallelic *Igf2* expression in HSCs (Fig. 3d–f). *Igf2* was similarly upregulated in bone marrow, including surrounding stromal cells, after maternal deletion of H19-DMR (Fig. 3g, h). However, as shown by reciprocal transplantation, an extrinsic increase of *Igf2* expression alone is not sufficient to cause the *mH19*^{ΔDMR/+} HSC phenotype. We next investigated whether the *Igf2* signalling pathway is activated in *mH19*^{ΔDMR/+} LT-HSCs. Binding of

Igf2 to *Igf1r* activates signalling, whereas *Igf2r* binding to *Igf2r* attenuates signalling¹⁴. *Igf2* and *Igf1r* amounts were gradually increased from LT-HSCs to MPPs (Fig. 3i, j). However, messenger RNA (mRNA), protein concentrations, and the number of *Igf1r*⁺ cells were significantly increased in *mH19*^{ΔDMR/+} LT-HSCs compared with controls (Fig. 3k–m), with no change in *Igf2r* expression (Fig. 3k). *Igf2*-*Igf1r* signalling is known to activate PI3K-Akt, which phosphorylates and inactivates FoxO3, a transcription factor that arrests the cell cycle^{15,16}. Inactive pFoxO3 was detected in only 15% of normal LT-HSCs but was substantially increased in ST-HSCs and MPPs (Fig. 3n, o); however, in *mH19*^{ΔDMR/+} pFoxO3 was detected in 75% of LT-HSCs. Our data indicate that H19-DMR deletion increased *Igf2*

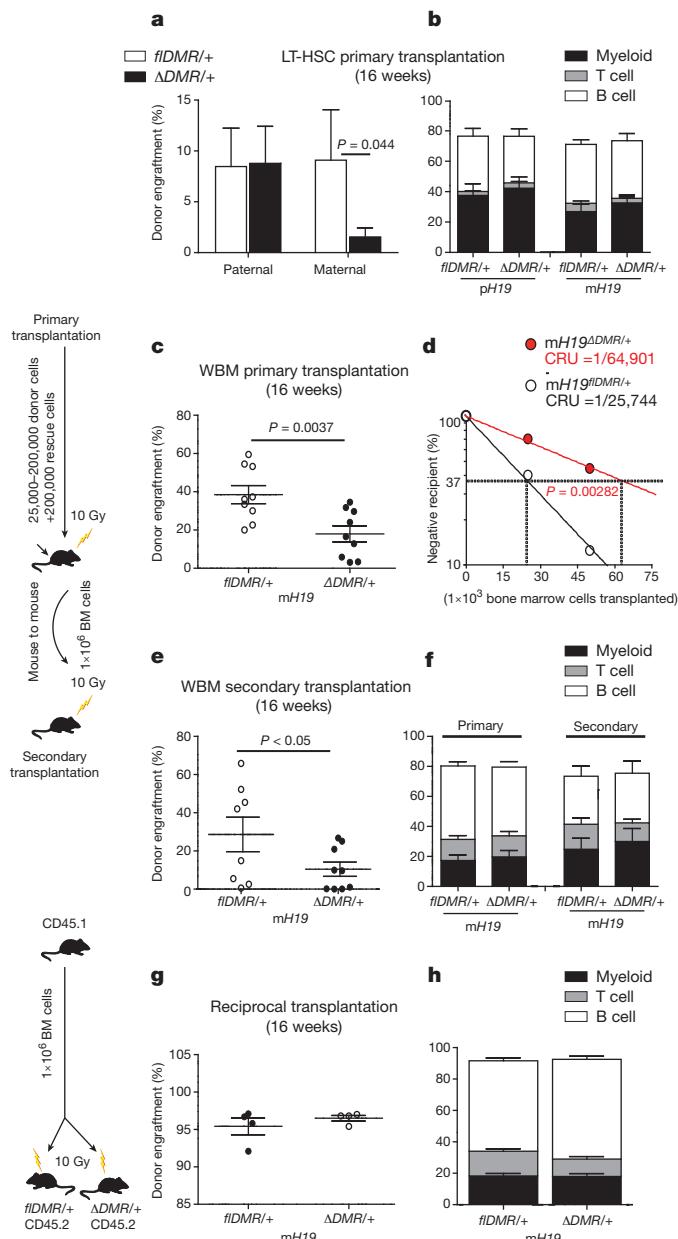


Figure 2 | Compromised HSC function in $mH19^{\Delta DMR/+}$ mice. Competitive re-population assay 16 weeks after transplantation in (a) sorted 100 LT-HSCs ($n = 10$), (c) primary transplant (test dosage of 2×10^5 bone marrow cells) ($n = 10$), (e) secondary transplant ($n = 10$), (g) reciprocal transplantation ($n = 4$). (d), Competitive re-population unit (CRU) content in each group of mice transplanted at each dose ($n = 60$ mice in total). Horizontal dotted line, 37% of recipient mice failed to engraft; vertical dotted lines, various CRU frequencies for each condition. Donor-derived lineage analysis after (b) primary transplantation, (f) secondary transplantation, (h) reciprocal transplantation. Error bars, s.e.m.

signalling, which released FoxO3-mediated suppression of HSC activation and proliferation.

In the placenta, *H19* functions as a precursor of miR-675, which in turn suppresses Igf1r⁶. We next investigated whether this regulation exists in adult HSCs. Expression of miR-675 was highest in LT-HSCs in the control mice but was substantially reduced in $mH19^{\Delta DMR/+}$ LT-HSCs (Fig. 4a). To explore the potential role of miR-675 in Igf1r regulation, we transplanted bone marrow cells overexpressing miR-675 into wild-type mice. Overexpression of miR-675 increased the percentage of quiescent CD34⁻ lineage⁻Sca-1⁺CKit⁺ (LSK) cells but did not significantly affect active CD34⁺ LSK cells (Fig. 4b). Western

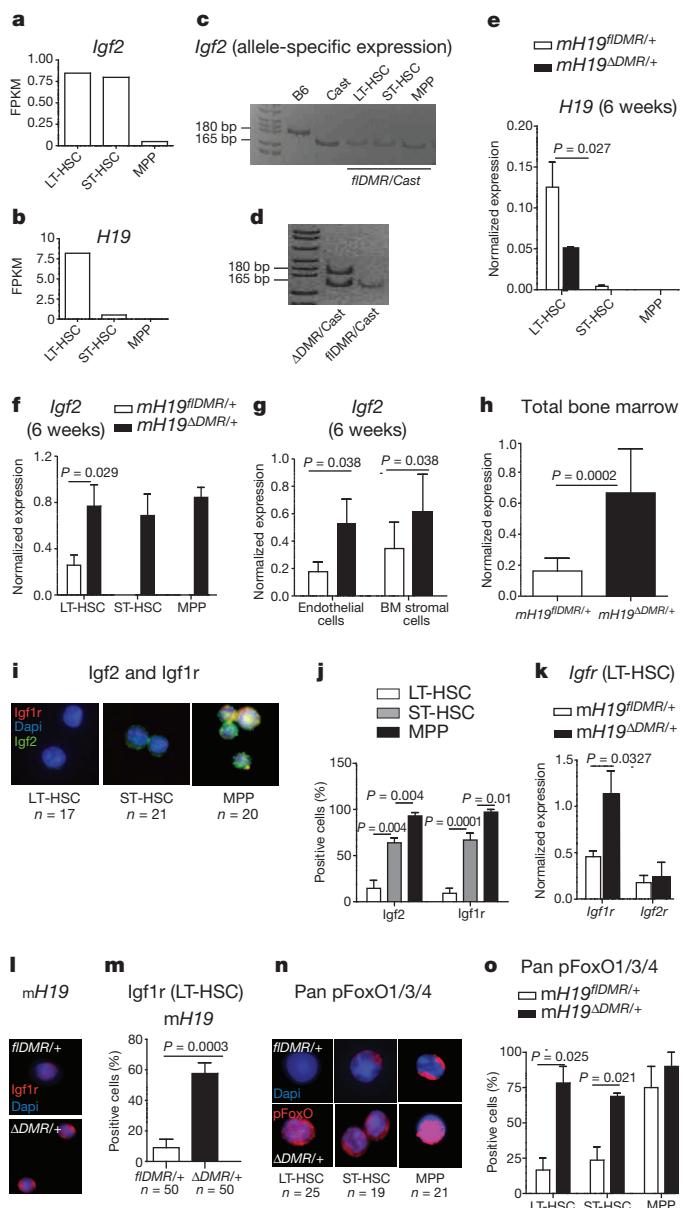


Figure 3 | Activation of Igf2-FoxO3 signalling in $mH19^{\Delta DMR/+}$ mice. **a, b**, RNA-seq analysis of *Igf2* and *H19* transcripts. FPKM, fragments per kilobase of exon per million fragments mapped. **c, d**, Allele-specific expression of the *Igf2* transcript. Quantitative (q)RT-PCR 6 weeks after pIpc induction ($n = 3$) for (e) *H19*, (f) *Igf2*, (g) stromal cells and (h) total bone marrow ($n = 5$). **i, j**, Single-cell Igf1r and Igf2 staining from wild-type bone marrow cells and its quantification ($n = 4$). **k, l**, Igf1r and Igf2r expression in sorted LT-HSCs. **m, n**, Igf1r immunostaining and its quantification ($n = 4$). **o**, Single-cell phospho-FoxO1/3/4 staining and its quantification ($n = 4$). Scale bar, 10 μ m. Error bars, s.e.m.

blot analysis showed a significant reduction of Igf1r by miR-675 compared with control (Fig. 4c). Furthermore, Igf1r amounts were lower in CD34⁻ LSK compared with CD34⁺ LSK cells in the control. However, miR-675 overexpression significantly decreased Igf1r (Fig. 4c-e). These data demonstrate that *H19*-derived miR-675 regulates Igf1r and the corresponding quiescent state in HSCs.

To confirm further that H19-DMR controls Igf2-Igf1r signalling, we crossed female $H19^{\Delta DMR/+}$ mice with male *Mx1-Cre:Igf1r^{fl/fl}* mice (Supplementary Fig. 4a)¹⁷. Although $mH19^{\Delta DMR/+}$ mutants (Fig. 3f, g) showed a decrease in LT-HSCs and an increase in ST-HSCs and MPPs, *Igf1r*^{-/-} mutants showed an increase in LT-HSCs and a decrease in MPPs. This indicates that Igf1r regulates the transitions from LT-HSCs to ST-HSCs and further to MPPs. Interestingly, $mH19^{\Delta DMR/+} Igf1r^{-/-}$

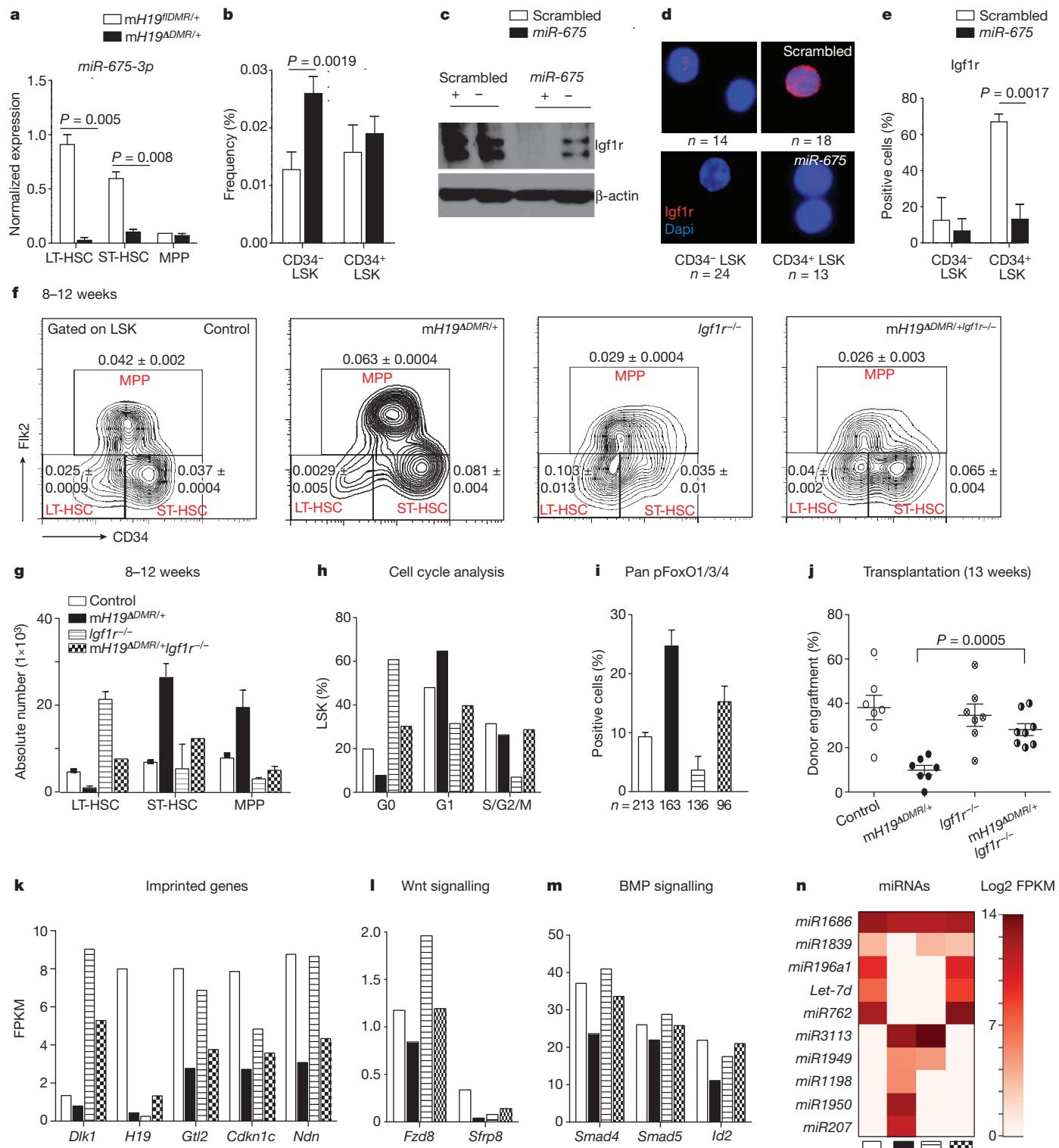


Figure 4 | Igf1r regulation and rescue by genetic blockage of Igf2–Igf1r signalling. a, miR-675 analysis by qRT-PCR ($n = 3$). b, Frequency of CD34⁺ versus CD34⁺ LSK cells 8 weeks after lentiviral infection. c, Immunoblot analysis of Igf1r in sorted bone marrow cells (positive and negative for green fluorescent protein (GFP)). d, e, Single-cell immunostaining of Igf1r 8 weeks

double mutants showed a partial restoration of LT-HSC frequency (Fig. 4f), whereas the transition from ST-HSCs to MPPs was still blocked (Fig. 4f, g). This indicates that Igf2–Igf1r signalling is partly responsible for the mH19^{ADMRI+} phenotype. An increase in ST-HSC frequency in double-mutant mice was probably due to blocked transition from ST-HSC to MPP by downregulation of Igf1r. Furthermore, cell cycle analysis

after lentiviral infection and its quantification ($n = 4$). f, Representative FACS plot with frequency ($n = 4$), (g) absolute number and (h) cell cycle analysis. i, Pan-pFoxO staining ($n = 3$). j, Donor engraftment 13 weeks after transplantation. RNA sequencing analysis in sorted stem cells: k, imprinted genes; l, m, Wnt and BMP signalling; n, heat map of miRNA expression. Error bars, s.e.m.

and pFoxO3 staining in double mutants showed partial rescue of the loss of quiescence phenotype (Fig. 4h, i).

To demonstrate phenotypic rescue functionally in mH19^{ADMRI+} Igf1r^{-/-} mice, we performed bone marrow transplantation assays. Although mH19^{ADMRI+} mutants had significantly reduced engraftment due to LT-HSC loss, engraftment of mH19^{ADMRI+} Igf1r^{-/-} bone

marrow cells increased to an amount between that of the mH19^{ADM}/+ and the Igf1r^{-/-} single mutants, indicating a partial functional rescue (Fig. 4j and Supplementary Fig. 4b). These results indicate that the maternal H19-DMR controls Igf2–Igf1r signalling which regulates HSC state; however, the partial rescue indicates that deletion of H19-DMR also affects other pathways required for LT-HSC maintenance. To investigate this possibility, we performed RNA-seq analysis of HSCs isolated from control, mH19^{ADM} single, Igf1r^{-/-} single and mH19^{ADM}/Igf1r^{-/-} double mutants. mH19^{ADM}/+ HSPCs showed widespread alterations in expression of imprinted genes in all three populations (Supplementary Figs 4c and 5). Genes involved in cell cycle arrest (*Cdkn1c*)^{18,19}, tumour suppression and stem cell maintenance (*Ndn*, *Gtl2*)^{20,21} were downregulated in LT-HSCs. However, Igf1r^{-/-} LT-HSCs largely maintained expression patterns similar to control, with high expression amounts of *Cdkn1c*, *Ndn*, *Gtl2* and *Dlk1*. The double mutants generally showed partial rescue of the alterations observed in single mutants, indicating either compensation in gene expression and/or existence of a proposed imprinted gene network^{22,23} (Fig. 4k and Supplementary Fig. 4c). Gene expression profiling of non-imprinted genes showed many overlapping downstream genes and microRNA (miRNAs) that were abnormally expressed in single mutants but partly rescued in double mutants (Fig. 4l–n and Supplementary Figs 6a–c and 7e). These included components of the Wnt and Tgf-β/BMP pathways such as *Smad4*, *Id2* and *Fzd8* (refs 24–27) as well as *Let-7*, which is known to repress cell proliferation²⁸ and Igf signalling²⁹ (Fig. 4l–n). Interestingly, H19-DMR potentially controls other miRNAs, small nucleolar RNAs (SnoRNAs) and genes (*Dusp26*, *p2rx2* and *Gpr63*) independent of the Igf2–Igf1r signalling (Supplementary Fig. 7a–f), with a known function for *Dusp26* in inhibiting cell proliferation³⁰. Taken together, our data show that maternal H19-DMR primarily restricts Igf2–Igf1r signalling, but also influences other genes and miRNAs involved in maintaining HSC quiescence.

By studying the H19-DMR locus in an allele-specific manner, we demonstrate that a specialized form of epigenetic control—genomic imprinting—is critical to the maintenance of adult stem cells. This is accomplished by maintaining LT-HSC quiescence, which can be attributed largely to Igf2–Igf1r-dependent signalling, but also to extra Igf2–Igf1r-independent effects on the regulation of cell cycle, proliferation and growth.

METHODS SUMMARY

Genotyping primers, flow assays and immunostaining were reported previously^{15,26}. Lentivirus-mediated transfection used Magnetofection™ and ViroMag R/L particles (OZ Biosciences). Microarray and RNA-seq analyses used Affymetrix MouseGenome430_2 arrays and Illumina TruSeq RNA Sample Prep Kit (catalogue number FC-122-1001) respectively, followed by sequencing on an Illumina HiSeq 2000. Statistical analyses were performed (pairwise comparisons were performed using a Student's *t*-test) in Prism 5 GraphPad.

Full Methods and any associated references are available in the online version of the paper.

Received 10 November 2011; accepted 15 May 2013.

Published online 17 July 2013.

- Bartolomei, M. S. Genomic imprinting: employing and avoiding epigenetic processes. *Genes Dev.* **23**, 2124–2133 (2009).
- Berg, J. S. et al. Imprinted genes that regulate early mammalian growth are coexpressed in somatic stem cells. *PLoS ONE* **6**, e26410 (2011).
- Ferrón, S. R. et al. Postnatal loss of *Dlk1* imprinting in stem cells and niche astrocytes regulates neurogenesis. *Nature* **475**, 381–385 (2011).
- Zacharek, S. J. et al. Lung stem cell self-renewal relies on BMI1-dependent control of expression at imprinted loci. *Cell Stem Cell* **9**, 272–281 (2011).
- DeChiara, T. M., Robertson, E. J. & Efstratiadis, A. Parental imprinting of the mouse insulin-like growth factor II gene. *Cell* **64**, 849–859 (1991).
- Keniry, A. et al. The H19 lincRNA is a developmental reservoir of miR-675 that suppresses growth and Igf1r. *Nature Cell Biol.* **14**, 659–665 (2012).

- Haug, J. S. et al. N-cadherin expression level distinguishes reserved versus primed states of hematopoietic stem cells. *Cell Stem Cell* **2**, 367–379 (2008).
- Yang, L. et al. Identification of Lin⁻Sca1⁺kit⁺CD34⁺Flt3⁻ short-term hematopoietic stem cells capable of rapidly reconstituting and rescuing myeloablated transplant recipients. *Blood* **105**, 2717–2723 (2005).
- Frost, J. M. & Moore, G. E. The importance of imprinting in the human placenta. *PLoS Genet.* **6**, e1001015 (2010).
- Hudson, Q. J., Kulinski, T. M., Huettner, S. P. & Barlow, D. P. Genomic imprinting mechanisms in embryonic and extraembryonic mouse tissues. *Heredity* **105**, 45–56 (2010).
- Thorvaldsen, J. L., Fedoriw, A. M., Nguyen, S. & Bartolomei, M. S. Developmental profile of H19 differentially methylated domain (DM) deletion alleles reveals multiple roles of the DM in regulating allelic expression and DNA methylation at the imprinted H19/Igf2 locus. *Mol. Cell. Biol.* **26**, 1245–1258 (2006).
- Lerner, C. & Harrison, D. E. 5-Fluorouracil spares hemopoietic stem cells responsible for long-term repopulation. *Exp. Hematol.* **18**, 114–118 (1990).
- Smith, F. M., Garfield, A. S. & Ward, A. Regulation of growth and metabolism by imprinted genes. *Cytogenet. Genome Res.* **113**, 279–291 (2006).
- Kang, H. M., Park, S. & Kim, H. Insulin-like growth factor 2 enhances insulinogenic differentiation of human eyelid adipose stem cells via the insulin receptor. *Cell Prolif.* **44**, 254–263 (2011).
- Zhang, J. et al. PTEN maintains haematopoietic stem cells and acts in lineage choice and leukaemia prevention. *Nature* **441**, 518–522 (2006).
- Tothova, Z. et al. FoxOs are critical mediators of hematopoietic stem cell resistance to physiologic oxidative stress. *Cell* **128**, 325–339 (2007).
- Klinakis, A. et al. Igf1r as a therapeutic target in a mouse model of basal-like breast cancer. *Proc. Natl. Acad. Sci. USA* **106**, 2359–2364 (2009).
- Zou, P. et al. p57(Kip2) and p27(Kip1) cooperate to maintain hematopoietic stem cell quiescence through interactions with Hsc70. *Cell Stem Cell* **9**, 247–261 (2011).
- Walkley, C. R., Shea, J. M., Sims, N. A., Purton, L. E. & Orkin, S. H. Rb regulates interactions between hematopoietic stem cells and their bone marrow microenvironment. *Cell* **129**, 1081–1095 (2007).
- Kubota, Y., Osawa, M., Jakt, L. M., Yoshikawa, K. & Nishikawa, S. Necdin restricts proliferation of hematopoietic stem cells during hematopoietic regeneration. *Blood* **114**, 4383–4392 (2009).
- Stadtfeld, M. et al. Aberrant silencing of imprinted genes on chromosome 12qF1 in mouse induced pluripotent stem cells. *Nature* **465**, 175–181 (2010).
- Zhao, Z. et al. Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. *Nature Genet.* **38**, 1341–1347 (2006).
- Varrault, A. et al. Zac1 regulates an imprinted gene network critically involved in the control of embryonic growth. *Dev. Cell* **11**, 711–722 (2006).
- Karlsson, G. et al. Smad4 is critical for self-renewal of hematopoietic stem cells. *J. Exp. Med.* **204**, 467–474 (2007).
- Sugimura, R. et al. Noncanonical wnt signaling maintains hematopoietic stem cells in the niche. *Cell* **150**, 351–365 (2012).
- Perry, J. M. et al. Cooperation between both Wnt/β-catenin and PTEN/PI3K/Akt signaling promotes primitive hematopoietic stem cell self-renewal and expansion. *Genes Dev.* **25**, 1928–1942 (2011).
- Dijke, P. & Heldin, C. H. *Smad Signal Transduction: Smads in Proliferation, Differentiation and Disease* (Springer, 2006).
- Johnson, C. D. et al. The let-7 microRNA represses cell proliferation pathways in human cells. *Cancer Res.* **67**, 7713–7722 (2007).
- Toledano, H., D'Alterio, C., Czech, B., Levine, E. & Jones, D. L. The let-7-Imp axis regulates ageing of the *Drosophila* testis stem-cell niche. *Nature* **485**, 605–610 (2012).
- Yu, W. et al. A novel amplification target, DUSP26, promotes anaplastic thyroid cancer cell growth by inhibiting p38 MAPK activity. *Oncogene* **26**, 1178–1187 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. Hembree, T. Johnson, H. Marshall, B. Lewis, D. Dukes, C. Semerad, J. Park and A. Box for technical support, and members of the Li laboratory for scientific discussion. We thank J. Lu and Y. Huang for communications about miR-675 and *Let-7*. We thank K. Tannen for editing. This work was supported by the Stowers Institute for Medical Research and by the Department of Biotechnology, Ministry of Science and Technology, Government of India, as an overseas associateship to A. Venkatraman. M. Bartolomei is supported by the National Institutes of Health (GM51279).

Author Contributions A.V. performed experiments, analysed data and wrote the manuscript. X.H. provided training, performed transplants and RNA-seq. F.T., J.T., M.C., L.P., X.Z., A.P., H.L., J.P., M.Z., J.H. and T.C. performed part of the experiments. M.B. contributed the mouse lines. L.L. directed the overall project and co-wrote the manuscript. All authors contributed to reading and editing the manuscript.

Author Information The microarray and RNA-seq data have been deposited in ArrayExpress under accession numbers E-MTAB-1644 and E-MTAB-1628, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to L. Li (lili@stowers.org).

METHODS

Animals. All mice used in this study were housed in the animal facility at the Stowers Institute for Medical Research and handled according to the guidelines of the Institute and the National Institutes of Health. All procedures were approved by the Institutional Animal Care and Use Committee of the Stowers Institute. *H19-DMR^{fIDMR/fIDMR}* mice on B6 background were provided by M. S. Bartolomei³¹. Conditional mutant *Igf1r^{f/+}* was provided by T. L. Clemens³². Interferon-inducible Mx1-Cre or tamoxifen-inducible Scl-Cre mouse strains were used to delete the floxed *H19-DMR* and *Igf1r*. For Mx1-Cre activation, 250 µg of pIpC was injected intraperitoneally every other day for 14 days at 5 weeks of age. For *Scl-CreER* activation, 2 mg of tamoxifen dissolved in 0.1 ml of corn oil was injected intraperitoneally every day for 5 days.

Single-cell HSC genotyping. Single CD34⁻ Flk2⁻ LSK cells were sorted into 96-well plates (one cell per well) containing 50 µl MethoCult complete media (M3434; Stem Cell Technologies) and incubated (37 °C, 5% CO₂) for 12 days. Individual colonies were collected separately, and DNA was purified using a QIAGENamp DNA Blood Kit (Qiagen). PCR genotyping of *H19* locus used G1, G5 and G7 primers described elsewhere³¹.

Flow cytometry. Phenotypic analyses of haematopoietic cells collected from bone marrow (femur and tibia) and peripheral blood were performed as described previously^{33,34}. Cell sorting and analysis were performed with a MoFlo (Dako) and/or CyAn ADP (Dako). We analysed data with FlowJo software (Ashland).

5FU treatment. Cohorts of *mH19^{ADMRI/+}* and *mH19^{fIDMR/+}* were injected with 5FU (Sigma-Aldrich) at 150 µg per g (body weight)³⁴ 4 weeks after pIpC induction. For one cycle, 5FU was injected once intravenously; for three cycles, 5FU was injected three times at 4-week intervals. Bone marrow cells were analysed 10 days after 5FU injections.

Cell cycle analysis. Cell cycle analysis of bone marrow LSKs was performed. Bone marrow cells (5×10^6) were stained for LSKs, fixed overnight at 4 °C in 4% paraformaldehyde and permeabilized with 0.2% Triton X-100. Cells were further stained with a BD Pharmingen™ FITC-conjugated-Mouse Anti-Human Ki67 Set (BD Pharmingen) according to manufacturer's instruction and 0.1 µg µl⁻¹ DAPI. This was followed by flow cytometric analysis with an InFlux Cell Sorter (BD Biosciences).

Transplantation studies. For competitive re-population assays, 2×10^5 bone marrow cells congenic with the host (CD45.1⁺) were included per mouse. One hundred sorted LT-HSCs from *mH19^{ADMRI/+}* or *mH19^{fIDMR/+}* were transplanted intravenously into lethally irradiated (10 Gy) Ptprc (CD45.1) recipient mice. Mice were placed on Baytril water 3 days before irradiation, which continued for 2 weeks after irradiation. Each transplanted group consisted of eight to ten recipients. Donor-derived engraftment was assayed every 4 weeks after transplantation by collection of peripheral blood, red blood cell lysis and staining of CD45.1 (recipient) versus CD45.2 (donor). Multi-lineage reconstitution was determined by CD3, B220 (for T and B lymphoid, respectively) and Gr1, Mac-1 (for myeloid) gating on donor (CD45.2⁺) cells. Limiting dilution experiments were performed by transplanting three doses (200,000, 100,000 and 25,000) of test samples ($n = 2$) from *mH19^{ADMRI/+}* or *mH19^{fIDMR/+}* (CD45.2) with a fixed number of 2×10^5 rescue cells (CD45.2) into groups of ten lethally irradiated (10 Gy) recipient mice (CD45.1). CRU frequency was determined with L-Calc software (Stem Cell Technologies) on the basis of Poisson statistics³⁵. The plot was made on the basis of the percentage of recipient mice containing 1% CD45.2⁺ cells in the peripheral blood at 16 weeks after transplantation versus the number of cells injected per mouse. For secondary transplantation, the original, primary transplant recipients were euthanized; bone marrow was collected from the femur and then transplanted mouse-to-mouse at a dosage of 1×10^6 cells per mouse. For reciprocal transplantation, wild-type congenic B6.SJL (CD45.1⁺) bone marrow cells (1×10^6 cells per recipient) were transplanted into lethally irradiated *Mx-1 Cre* induced *mH19^{fIDMR/+}* and *mH19^{ADMRI/+}* (CD45.2⁺) recipients. Complete donor cell engraftment by wild-type CD45.1⁺ cells was confirmed by flow cytometric analysis. For rescue transplants, 2×10^5 (CD45.2) bone marrow cells from the *mH19^{fIDMR/+}*, *Igf1r^{-/-}* and *mH19^{ADMRI/+}* *Igf1r^{-/-}* mutants and the controls along with 2×10^5 (CD45.1⁺) Ptprc bone marrow cells were transplanted into CD45.1 recipient mice. Complete donor cell engraftment by wild-type CD45.1⁺ cells was confirmed by flow cytometric analysis.

Lentivirus infection. Mice were treated with 150 µg per g (body weight) of 5FU to activate and enrich for HSPCs³⁵. Four days later, bone marrow was collected and cultured overnight in ST media and transduced by Magnetofection™ using ViroMag R/L particles according to the manufacturer's protocol (OZ Biosciences). Transplantation experiments conducted in the knockdown model were done with 300,000 unsorted infected bone marrow cells (CD45.2). The cells were transplanted into each lethally irradiated (10 Gy) Ptprc (CD45.1). Eight weeks after engraftment, bone marrow cells that were double positive for GFP and CD45.2 were sorted for CD34⁻ LSK and CD34⁺ LSK.

Lentivirus construction. The pSicoR-EF1α promoter-IRES-EGFP lentiviral construct was provided by T. Xie.

mir-675: forward, AGCGTGCAGCCCAAGGGACTGGTCGCGAAAGGGCCA CAGTGGACTTGGTACACTGTATGCCCTAACGCTCAGTCCCTGGTC TGGCA; reverse, GGCGATGCCAGACCCAGGGACTGAGCGGTTAGGGCAT ACAGTGACCAAGTCCACTGTGGCCCTTCCGACCAGTCCCTGGC CCGCA.

IGF2 shRNA: forward, AGCGCGCCCAAATTGATTGGCTCAAATAGTGA AGCCACAGATGTATTAGGCCAATCAAATTGGTCA; reverse, GGCGAT GACCAAATTGATTGGCTCTAAATACATCTGTGGCTTCACTATTAGA GCGAACATCAATTGGCG.

Allele-specific Igf2 expression. Male Cast mice were crossed with female fIDMR or ADMR. Heterozygous progeny at single nucleotide polymorphisms differ between the two strains. In mice that inherited the Cast allele paternally and fIDMR maternally, LT-HSCs, ST-HSCs and MPPs were sorted from total bone marrow cells. RNA was extracted and DNase treated using RQ1 RNase-Free DNase according to the manufacturer's instructions (Promega). This RNA was reverse transcribed in the presence of SuperScriptIII Reverse Transcriptase (SSIII) using *Igf2*-specific primer Igf2-20r (5'-gggttgttttagccatcaa-3') according to the manufacturer's instructions (Invitrogen); simultaneously, equal concentrations of RNA were identically treated in the absence of SSIII (for minus (-) RT). Equal volumes of RT and -RT were amplified using 0.5 µM of primers Igf2-18f (5'-atctgtaccccttgcaggcagg-3') and Igf2-20r and Go-Taq Green Master Mix (Promega) using the following PCR conditions: 94 °C for 2 min for one cycle; 94 °C for 15 s, 58 °C for 15 s, and 72 °C for 20 s for 43 cycles. No product was detected in -RT samples. Amplified *Igf2* was digested with *Mlu*CI (NEB) and the paternal Cast product (165 base pairs (bp)) and the maternal B6 product (180 bp) were resolved on a 15% polyacrylamide gel similar to methods described earlier³¹. **Microarray.** RNA was extracted by conventional TRIzol® method from sorted LT-HSCs, ST-HSCs and MPPs (10,000 cells each)³⁶. Samples were analysed with Affymetrix MouseGenome430_2 arrays and scanned with a GeneChip Scanner 3000 7G using GeneChip Fluidics Station 450 and GeneChip Operating Software (GCOS 1.4). Heat map data represent the fold change between two populations from at least three independent biological samples. Three samples were run on Affymetrix Mouse 430.2 arrays in triplicate, using the standard Affymetrix protocols. CEL files were read into the R software environment (<http://www.r-project.org/>) and normalized with RMA³⁷⁻³⁹. Normalized data were fitted with a linear model using the limma package⁴⁰ and three contrasts were fitted: CD34p/CD34n, FLK2p/CD34p, FLK2p/CD34n. Probes that were significant for at least one contrast (BH adj *P* < 0.05) went to further analysis.

A list of imprinted genes was taken from the catalogue of imprinting genes at <http://igc.otago.ac.nz/1101Summary-table.pdf>. Names were matched to MGI and Ensembl 63 genes and then converted to probeset identifications. Of 125 input genes, 86 could be mapped to probesets, and, of these, 23 were significant. Sample expression coefficients per probeset were averaged together by gene. Expression amounts varied widely, from 4 to 14 in log₂ scale, which obscured the regulatory trend across samples during clustering. We constructed a heat map to show only the trending of expression, not the magnitude, by subtracting the mean from each row and dividing by the standard deviation. Thus, the heat map scale shows expression z-scores. Row ordering reflects hierarchical clustering, average linkage, using Pearson dissimilarity for distance. Microarray data have been deposited in ArrayExpress under accession number E-MTAB-1644.

qRT-PCR. Total RNA (2–50 ng) was extracted from sorted LT-HSCs, ST-HSCs and MPPs directly into TRIzol®. This was followed by DNase I treatment (Ambion) and reverse transcription with a high-capacity complementary DNA (cDNA) reverse transcription kit (Applied Biosystems). cDNA was pre-amplified by TaqMan PreAmplification master mix (Applied Biosystems) according to the manufacturer's instructions. TaqMan gene expression assays (Applied Biosystems) were performed on triplicate samples with a 7500 Real-Time cycler (Applied Biosystems). Data were normalized relative to *Gapdh* and *Hprt1*. For mi-R-675-3p assay, extracted RNA were reverse transcribed using TaqMan miRNA reverse transcription kit (Applied Biosystems). TaqMan pre-amplification and TaqMan gene expression assay were performed according to the manufacturer's instructions. All qRT-PCR was performed using TaqMan probes.

Immunostaining. Immunostaining was performed as described previously³³. For immunostaining of sorted cells, cells were sorted onto lysine-coated slides, fixed with chilled methanol for 10 min, followed by blocking and staining with primary antibody⁴¹. The following primary antibodies were used: chicken anti *Igf1r* (Abcam, 1:100), rabbit anti *Igf2* (Abcam, 1:100), rabbit anti FoxO3a (1:100), rabbit anti Foxo1/3/4-Pan and phosphor_Thr24/32 (Assay biotech, 1:50).

RNA-seq. The RNA-sequencing library was prepared from approximately 200 ng of total RNA (*mH19^{ADMRI/+}* *Igf1r^{-/-}*, *mH19^{ADMRI/+}* *Igf1r^{f/-}* and *mH19^{fIDMR/+}*) for each sample using an Illumina TruSeq RNA Sample Prep Kit (catalogue number

FC-122-1001). The fragment size in the generated library ranged from 220 to 500 bp with a pick at 280 bp. A total of 10 fmol library fragments were loaded to cBot to generate clusters, followed by sequencing on an Illumina HiSeq 2000 to produce 10 million to 30 million paired-end 100 bp reads per sample. Reads were trimmed to 70 bp owing to quality and aligned to mm9 with Tophat 1.3.1 (ref. 42)/Bowtie 0.12.7 (ref. 43), using the Ensembl 63 GTF file for gene models. Parameters were —g 1, mate-inner-dist 200, mate-std-dev 70, segment-length 35, segment-mismatches 2; this allowed for four mismatches per read (two per read half) and unique alignments only.

Gene expression was quantified using Cufflinks 1.0.3 (ref. 44). We chose any imprinted genes with an absolute log-fold change of 1.3. A total of 38 imprinted genes were selected this way, 32 having measurements on both samples and six having measurements in only one sample. We heat mapped the genes found in both samples using FPKMs only. For the genes found in both, the range of expression was skewed enough to make visualization by heat map difficult, so we created a row-normalized heat map as with the microarray data. RNA-seq data have been deposited in ArrayExpress under accession number E-MTAB-1628.

31. Thorvaldsen, J. L., Fedoriw, A. M., Nguyen, S. & Bartolomei, M. S. Developmental profile of H19 differentially methylated domain (DMD) deletion alleles reveals multiple roles of the DMD in regulating allelic expression and DNA methylation at the imprinted H19/Igf2 locus. *Mol. Cell. Biol.* **26**, 1245–1258 (2006).
32. Dietrich, P., Dragatsis, I., Xuan, S., Zeitlin, S. & Efstratiadis, A. Conditional mutagenesis in mice with heat shock promoter-driven cre transgenes. *Mamm. Genome* **11**, 196–205 (2000).
33. Zhang, J. *et al.* Identification of the haematopoietic stem cell niche and control of the niche size. *Nature* **425**, 836–841 (2003).
34. Haug, J. S. *et al.* N-cadherin expression level distinguishes reserved versus primed states of hematopoietic stem cells. *Cell Stem Cell* **2**, 367–379 (2008).
35. Miller, C. L. & Eaves, C. J. Expansion in vitro of adult murine hematopoietic stem cells with transplantable lympho-myeloid reconstituting ability. *Proc. Natl. Acad. Sci. USA* **94**, 13648–13653 (1997).
36. Akashi, K. *et al.* Transcriptional accessibility for genes of multiple tissues and hematopoietic lineages is hierarchically controlled during early hematopoiesis. *Blood* **101**, 383–389 (2003).
37. Bolstad, B. M., Irizarry, R. A., Astrand, M. & Speed, T. P. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. *Bioinformatics* **19**, 185–193 (2003).
38. Irizarry, R. A. *et al.* Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
39. Irizarry, R. A. *et al.* Summaries of Affymetrix GeneChip probe level data. *Nucleic Acids Res.* **31**, e15 (2003).
40. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
41. Ema, H. *et al.* Adult mouse hematopoietic stem cells: purification and single-cell assays. *Nature Protocols* **1**, 2979–2987 (2006).
42. Trapnell, C., Pachter, L. & Salzberg, S. L. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics* **25**, 1105–1111 (2009).
43. Langmead, B., Trapnell, C., Pop, M. & Salzberg, S. L. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* **10**, R25 (2009).
44. Trapnell, C. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nature Biotechnol.* **28**, 511–515 (2010).

PAAR-repeat proteins sharpen and diversify the type VI secretion system spike

Mikhail M. Shneider^{1,2*}, Sergey A. Buth^{1*}, Brian T. Ho³, Marek Basler³, John J. Mekalanos³ & Petr G. Leiman¹

The bacterial type VI secretion system (T6SS) is a large multicomponent, dynamic macromolecular machine that has an important role in the ecology of many Gram-negative bacteria. T6SS is responsible for translocation of a wide range of toxic effector molecules, allowing predatory cells to kill both prokaryotic as well as eukaryotic prey cells^{1–5}. The T6SS organelle is functionally analogous to contractile tails of bacteriophages and is thought to attack cells by initially penetrating them with a trimeric protein complex called the VgrG spike^{6,7}. Neither the exact protein composition of the T6SS organelle nor the mechanisms of effector selection and delivery are known. Here we report that proteins from the PAAR (proline-alanine-alanine-arginine) repeat superfamily form a sharp conical extension on the VgrG spike, which is further involved in attaching effector domains to the spike. The crystal structures of two PAAR-repeat proteins bound to VgrG-like partners show that these proteins sharpen the tip of the T6SS spike complex. We demonstrate that PAAR proteins are essential for T6SS-mediated secretion and target cell killing by *Vibrio cholerae* and *Acinetobacter baylyi*. Our results indicate a new model of the T6SS organelle in which the VgrG-PAAR spike complex is decorated with multiple effectors that are delivered simultaneously into target cells in a single contraction-driven translocation event.

The T6SS organelle is a 1,000-nm long tubular structure consisting of an inner tube made of multiple copies of the Hcp protein and an external contractile sheath composed of VipA and VipB proteins (also known as TssB/TssC)³. It has been visualized in two conformations (extended and contracted), which are both attached to the cell envelope by means of a baseplate complex³. Rapid contraction of the sheath results in translocation of the inner tube out of the predator cell and into the prey cell³. The trimeric VgrG spike protein is positioned at the end of the tube⁷. It is believed to penetrate the prey cell with its needle-shaped carboxy-terminal β-helical domain. VgrG proteins can contain additional C-terminal domains that act as effectors⁶ and may also bind effectors through undefined mechanisms⁸.

Because VgrG proteins are orthologous to the central baseplate spikes of bacteriophages with contractile tails, we reasoned that additional structural components present in certain phage spikes might have corresponding orthologous components in T6SS. The cryoEM reconstruction of phage T4 baseplate shows that an unknown protein with a molecular weight between 7 and 23 kilodaltons (kDa) binds to the tip of the β-helical domain of the central spike protein gp5 (ref. 9). We analysed all known genes encoding small proteins in phage genomes with gp5-like spikes and compared them to T6SS genes. Proteins containing the PAAR-repeat motif were strongly represented in this group with gp5.4 being the corresponding protein of T4 phage. Furthermore, genes encoding proteins with PAAR motifs were frequently found immediately downstream from *vgrG*-like genes suggesting that the two are genetically linked⁷. Therefore, we devised a strategy to test the hypothesis that these PAAR proteins were binding to the tip of gp5 and VgrG proteins.

Careful examination of VgrG sequences showed that a β-structural repeat, which is presumed to be responsible for β-helix formation^{6,10},

either extends to the very C terminus of the protein or terminates with a glycine/serine-rich stretch (Extended Data Fig. 1). We surmised that the glycine/serine-rich stretch bends the polypeptide chain away from the β-helix without disturbing its tip, and that all VgrG β-helices have blunt ends resembling that of T4 gp5 β-helix. The X-ray crystal structure of one full-length VgrG trimer supports this conclusion (Sycheva L. V., Shneider M. M., Basler M., Ho B., Mekalanos J. J. & Leiman P. G., manuscript in preparation). Thus, we further speculated that the binding site for a PAAR protein was the blunt end of the β-helix and designed experiments to test this idea.

Because of the solubility problems of most tested VgrG and PAAR proteins, we used a soluble fragment of T4 gp5 β-helix (residues 484–575), which is known to fold into a stable native-like trimeric structure¹⁰, as a platform for creating β-helices mimicking the blunt end of various VgrG spikes. We replaced the tip of gp5 β-helix (the last two β-strands) with an equivalent fragment from several putative β-helices of VgrGs from *Escherichia coli*, *Vibrio cholerae* and *Pseudomonas aeruginosa* (Extended Data Fig. 1). These gp5–VgrG chimaeras (hereafter referred to as gp5(VgrG)) were then co-expressed with various PAAR proteins, in *cis* and in *trans*, in different combinations (Extended Data Fig. 1). Surprisingly, the *V. cholerae* VCA0105 and *E. coli* c1882 PAAR proteins were found to bind not only to the ends of their endogenous VgrG proteins but also to the non-mutated wild-type end of the gp5 β-helix as well.

The gp5(c1883)–c1882 and gp5(VCA0018)–VCA0105 complexes were purified to homogeneity and crystallized (Extended Data Table 1). The structures were solved by molecular replacement¹¹ using the corresponding gp5 β-helix fragment as a search model¹⁰. In both complexes, a single chain of the PAAR protein folds into a symmetrical cone-shaped structure with a sharp tip and a triangular base fully occupying the blunt end of the β-helix (Fig. 1). The cone contains nine short β-strands, three of which create its base and participate in binding to the VgrG β-helix and six others form three β-hairpins that point towards the vertex of the cone, but have different lengths. The PAAR proteins interact with the VgrG β-helix by means of a virtually flat hydrophobic patch and 14 or 16 hydrogen bonds for c1882 or VCA0105, respectively (Extended Data Fig. 2). In both proteins, 12 hydrogen bonds between the main chain atoms of the VgrG β-helix and those of the PAAR domain form a perfect triangle surrounding the central hydrophobic patch creating a unique binding platform (Extended Data Fig. 3). PISA software¹² shows that ~16.5% of the PAAR protein surface is buried in this interface and the free energy of interaction between the VgrG tip and c1882 or VCA0105 PAAR proteins is $-5.5 \text{ kcal mol}^{-1}$ or $-3.4 \text{ kcal mol}^{-1}$, respectively.

The fold of the PAAR protein is stabilized by a zinc atom positioned close to the cone's vertex (Fig. 1d, Extended Data Fig. 4). The Zn binding site consists of three histidines and one cysteine—H14, H46, H54 and C81 in VCA0105—that are very well conserved in close homologues (Extended Data Fig. 5a). These residues are replaced with similar or complementary metal-binding residues (arginines, lysines

¹École Polytechnique Fédérale de Lausanne (EPFL), BSP-415, 1015 Lausanne, Switzerland. ²Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Laboratory of Molecular Bioengineering, 16/10 Miklukho-Maklaya Street, 117997 Moscow, Russia. ³Department of Microbiology and Immunobiology, Harvard Medical School, 77 Avenue Louis Pasteur, Boston, Massachusetts 02115, USA.

*These authors contributed equally to this work.

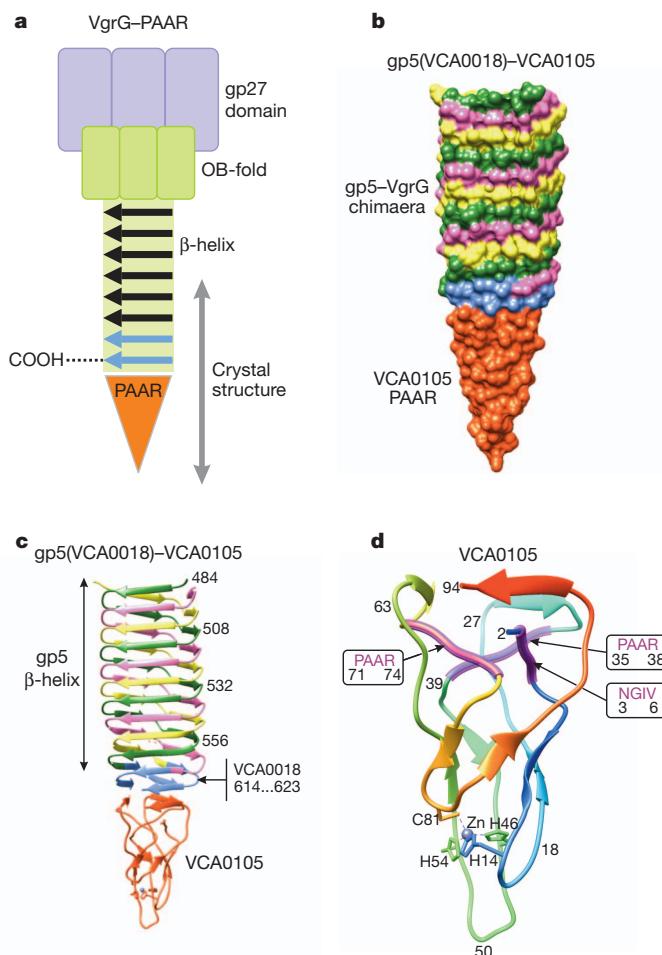


Figure 1 | Crystal structure of the VCA0105 PAAR-repeat protein bound to its VgrG-like partner. **a**, Schematic representation of the conserved domains comprising the VgrG-PAAR complex. The last strands of the β -helix that form the PAAR binding site are in light blue. The grey arrow shows the fragment roughly corresponding to the crystal structure. **b**, Molecular surface representation of the gp5(VCA0018)-VCA0105 complex crystal structure. Each protein chain is labelled with its own colour. **c**, Ribbon diagram of the gp5(VCA0018)-VCA0105 complex. **d**, The polypeptide chain of the VCA0105 PAAR protein is coloured in rainbow colours with the N terminus in blue and C terminus in red. Residues responsible for Zn binding are labelled. PAAR motifs are highlighted with thick purple coils. The sequence and residue numbers of the starting and ending position for each of the three PAAR motifs are boxed.

PAAR motif meet and intertwine (Fig. 1d, Extended Data Fig. 5b, c). Hydrophobic interactions and buried main chain hydrogen bonds mediate the interaction of the three PAAR motifs. (Extended Data Fig. 5b, c). The distance between the C_α atoms of the third residue of the three PAAR motifs in both proteins is only 5.3 ± 0.1 Å (Extended Data Fig. 5b, c). c1882 and VCA0105 show 61.3% sequence identity and as a consequence the two structures are very similar with a root mean square deviation (r.m.s.d.) of 0.53 Å between all the 94 C_α atoms comprising the backbone (Extended Data Fig. 6).

Given such a critical location of PAAR proteins within the T6SS organelle and their high structural conservation, we speculated that inactivation of PAAR genes would probably interfere with T6SS functions including protein secretion and prey cell killing. Because some T6SS⁺ organisms have many gene products predicted to carry the PAAR motif, we focused on *Vibrio cholerae* strain 2740-80 and *Acinetobacter baylyi* ADP1, which each encode only two and three PAAR proteins, respectively. In *A. baylyi* inactivation of all three PAAR genes (*aciad0051*, *aciad0052*, *aciad2681*) led to an ~90% reduction in Hcp secretion (Extended Data Fig. 7) and at least a 10,000-fold decrease in T6SS-dependent killing of *E. coli* (Fig. 2a, c). Similarly, inactivation of both genes (*vca0105* and *vca0284*) that encode PAAR proteins in *V. cholerae* 2740-80 resulted in a ~100-fold decrease in T6SS-dependent killing of *E. coli* and ~70% reduction in Hcp secretion (Fig. 2b, c; Extended Data Fig. 7). Remarkably, in both bacterial species, mutants inactivated in single PAAR genes showed no or only a modest defect in the functionality of their T6SS apparatus (Fig. 2) indicating that PAAR proteins within a species are interchangeable. We further attempted to complement these double and triple PAAR mutants with heterologous PAAR genes but these experiments were inconclusive (Extended Data Fig. 8). Thus, PAAR proteins may be very specific for their cognate VgrG proteins in the context of the functional T6SS organelle.

The dramatic reduction in Hcp secretion in the triple PAAR gene knockout mutant of *A. baylyi* suggests that PAAR proteins participate

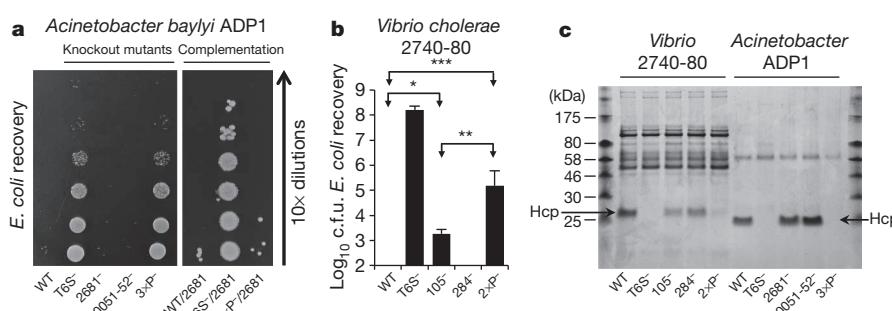


Figure 2 | PAAR proteins are required for full functionality of the T6SS in *Vibrio cholerae* and *Acinetobacter baylyi*. **a**, Recovery of viable *E. coli* MG1655 after co-incubation with *A. baylyi* ADP1 (wild type, WT) and its T6SS and PAAR genes knockout mutants. The following genes were inactivated in the mutants shown: T6S⁻, *aciad2688* to *aciad2694*; 2681⁻, *aciad2681*; 0051-52⁻, both *aciad0051* and *aciad0052*; 3 X P⁻, all three PAAR genes *aciad0051*, *aciad0052* and *aciad2681*. Leaky, basal expression of the *aciad2681* PAAR gene from plasmid pMMB67EH²⁶, labelled as p2681, restores the killing defect in the triple PAAR mutant (right panel). **b**, Recovery of *E. coli* MG1655 colony forming units (c.f.u.) after co-incubation with *V. cholerae* 2740-80 and its T6SS and PAAR gene knockout mutants, which are labelled as follows: T6S⁻, *vipA*⁻, *vca0105*; 284⁻, *vca0284*; 2 X P⁻, *vca0105* and *vca0284*. Symbols *, ** and *** indicate deviations from the wild type with P values of 6×10^{-3} , 8×10^{-3} and 5×10^{-7} , respectively, for a sample size of 8. Error bars represent one standard deviation. **c**, SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis assay) for T6SS-dependent secretion of Hcp proteins by the parental strains and T6SS and PAAR genes knockout mutants. Panels **a** and **c** show one out of three experiments with similar outcomes.

forming units (c.f.u.) after co-incubation with *V. cholerae* 2740-80 and its T6SS and PAAR gene knockout mutants, which are labelled as follows: T6S⁻, *vipA*⁻, *vca0105*; 284⁻, *vca0284*; 2 X P⁻, *vca0105* and *vca0284*. Symbols *, ** and *** indicate deviations from the wild type with P values of 6×10^{-3} , 8×10^{-3} and 5×10^{-7} , respectively, for a sample size of 8. Error bars represent one standard deviation. **c**, SDS-PAGE (sodium dodecyl sulphate-polyacrylamide gel electrophoresis assay) for T6SS-dependent secretion of Hcp proteins by the parental strains and T6SS and PAAR genes knockout mutants. Panels **a** and **c** show one out of three experiments with similar outcomes.

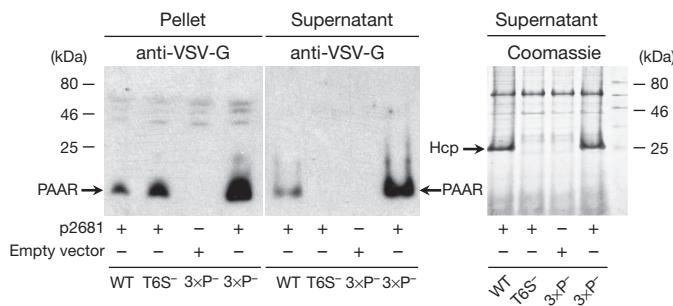


Figure 3 | The VSV-G epitope-tagged PAAR protein ACIAD2681 is secreted by *A. baylyi* ADP1. T6SS-dependent secretion of VSV-G epitope-tagged ACIAD2681 expressed from plasmid pMMB67EH (left panels). VSV-G epitope-tagged ACIAD2681 fully restores the Hcp secretion defect of the triple knockout PAAR mutant of *A. baylyi* ADP1 (right panel). The mutants are labelled as in Fig. 2. A representative of three identical experiments was chosen for each panel.

in the assembly of the T6SS complex by either nucleating the folding of VgrG trimers or regulating their incorporation into the T6SS organelle. Similar to the T4 phage system in which mutants lacking gp5 (VgrG orthologue) are tailless¹³, any disruption in VgrG trimeric assembly will probably block the assembly of the T6SS organelle. Alternatively, PAAR proteins may be important for another function of the apparatus such as translocation of the VgrG spike through the predator outer membrane during a sheath contraction event.

The PAAR-repeat proteins form a diverse superfamily called CL15808 in the conserved domains (CDD) database¹⁴ that contains three families PF05488, COG4104 and PF13665 (or DUF4150). The first two families are similar and describe PAAR domains that are on average ~95 residues long, whereas the PF13665 family is somewhat more distant and its typical members contain ~130 amino acids. The crystal structures reported here include representatives of the PF05488 family. Hundreds of hypothetical proteins in the database contain PAAR domains that are extended both N- and C-terminally by domains with various predicted functions (Extended Data Fig. 9). The crystal structure shows that the termini of the PAAR domain are open to solution and thus can be extended without distorting the VgrG binding site (Fig. 1, Extended Data Fig. 6).

The C-terminal domains of many hypothetical PAAR proteins are predicted to have various enzymatic activities that are toxic for prokaryotic and eukaryotic cells (Extended Data Fig. 9). Very similar putative effector domains can be found fused to the C termini of VgrG proteins (for example, VIP2 ADP-ribosyl transferase). Binding of these larger PAAR proteins to the tip of VgrG spikes would decorate the T6SS spike with a great variety of effector domains. To test this hypothesis, we expressed a VSV-G epitope-tagged version of the ACIAD2681 PAAR protein in the wild-type *A. baylyi* ADP1 and in its triple PAAR gene knockout mutant (Fig. 3). The epitope-tagged protein was secreted by T6SS and fully restored T6SS-mediated killing of *E. coli* in the triple PAAR gene knockout strain (Figs 2a and 3), suggesting that other effector proteins fused to PAAR domains will probably also be targeted for secretion by binding the tip of VgrG trimers.

VCA0284, the larger of the two *V. cholerae* V52 PAAR proteins, carries a transthyretin domain (TTR) at its C terminus, which is a very common architecture of PAAR proteins (Extended Data Fig. 9). TTR is an immunoglobulin-like domain that is known to form oligomeric structures in which these domains interact with each other or with other partners¹⁵. Thus, PAAR-associated TTR domains may act as adapters to further decorate the VgrG tip with effectors displaying TTR domains or serve to bind the spike to other TTR domain-containing proteins such as the TssJ/SciN lipoprotein¹⁶, an outer membrane structural component of the T6SS ‘baseplate’³.

Besides being in T6SS gene clusters, many PAAR genes are frequently found downstream of *vgrG* genes⁶, which in turn, are often

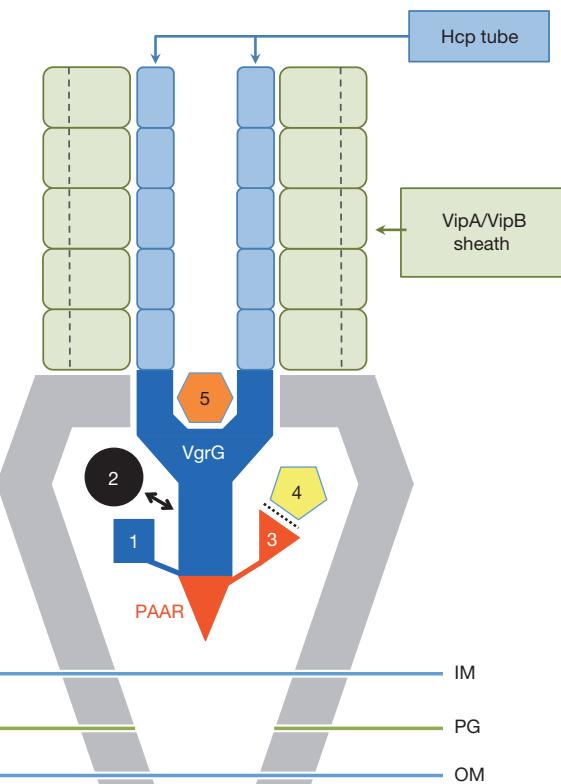


Figure 4 | Multiple effector translocation VgrG (MERV) model for the organization of the T6SS central spike/baseplate. Effectors are predicted to be loaded onto the spike complex by five distinct mechanisms: 1) C-terminal extensions of the VgrG spike; 2) non-covalent binding to the VgrG spike; 3) N- or C-terminal extensions of the PAAR protein; 4) non-covalent binding to the PAAR protein or its extension domains; 5) incorporation into the cavity formed by the gp27 domain of VgrG. A single T6SS sheath contraction event translocates the VgrG spike with all of its cargo proteins into a nearby target cell. Other proteins making up the T6SS ‘baseplate’ (grey colour) are not labelled but presumably reside within or attached to the inner and outer membranes and peptidoglycan layer (IM, OM, and PG, respectively).

encoded by Rhs elements (accessory sequences that have the hallmarks of horizontally acquired DNA, Extended Data Fig. 9). A recent report¹⁷ shows that two different nucleases encoded by the *rhsA* and *rhsB* genes of *Dickeya dadantii* are delivered into target cells in a VgrG-dependent process, suggesting that the T6SS locus of this organism mediates translocation of these effectors. Interestingly, both RhsA and RhsB proteins contain PAAR repeat regions and other features that categorize them as Class 7 PAAR domain architecture proteins (Extended Data Fig. 9b). Thus, our working model (Fig. 4) predicts that these T6SS nuclease effectors should bind to the tip of VgrG trimers through their PAAR repeat domains and in this way be targeted for secretion and translocation into prey cells.

The structural, functional and bioinformatic findings summarized above allow us to make several conclusions and predictions. Because a PAAR-repeat protein caps the end of the β-helix of a VgrG spike, it is in fact the piercing tip that is responsible for the initial event of creating an opening in the target cell envelope. Furthermore, because the crystal structure demonstrates that the canonical PAAR-repeat domain can be extended N- or C-terminally without distorting its structure or its VgrG β-helix binding site, we predict that large PAAR proteins carrying effector domains will probably also bind to VgrG spikes and be translocated into target cells by the T6SS organelle. Considering the findings reported here and other published data, we propose that there are five mechanisms by which effectors can be incorporated into the T6SS spike complex (Fig. 4). Three of them: 1) C-terminal extensions of the VgrG spike⁶; 2) binding surface features on the VgrG protein^{8,18};

and 3) N- or C-terminal extensions of the PAAR protein¹⁷; are supported by direct or indirect experimental evidence. Two others: 4) binding surface features or additional domains (for example, the TTR domain) on PAAR proteins and 5) incorporation into the cavity formed by the gp27 domain of VgrG, remain speculative. Thus, the T6SS machine may be capable of delivering a multifunctional ‘cargo’ or multiple effector translocation VgrG (MERV) spike into the target prey cell in a single molecular translocation event driven by T6SS sheath contraction.

METHODS SUMMARY

Structure determination of gp5(VgrG)-PAAR complexes. The gp5 fragment containing residues 484–575 and its variants with modified C termini mimicking VgrG were cloned into a derivative of the pET-23a expression vector (Novagen) with a TEV-cleavable N-terminal polyhistidine tag. For expression in *cis*, PAAR genes were cloned downstream from the gp5 fragment with a spacer containing the AGGAGG ribosome-binding site. For expression in *trans*, PAAR genes were cloned into a modified pACYCDuet-1 expression vector (Novagen), retaining its wild-type N and C termini. The complexes were purified by metal-affinity, ion exchange and size exclusion chromatographies. Conditions yielding diffraction quality crystals were refined from the initial hits obtained with the help of commercial crystallization screens (Jena Bioscience). The structures were solved by molecular replacement with the help of PHASER¹⁹ using the corresponding fragment of the gp5 β-helix as a search model (Protein Data Bank (PDB) accession ID number 1K28)¹⁰. Crystallographic refinement was performed with PHENIX²⁰, REFMAC5²¹ and COOT²².

Bioinformatics. 1,353 putative orthologues of PAAR repeat proteins were identified with the help of HMMER²³. Phage proteins and orphan genes were excluded from this analysis. Functional assignment was based on high confidence hits obtained with the help of HHpred²⁴, HMMER²³, CDD¹⁴ and CDART²⁵.

Bacterial killing assays. Bacterial competitions and statistical analysis involving *V. cholerae* 2740-80, *A. baylyi* ADP1 and *E. coli* MG1655 were performed as described previously⁵. The plasmid pMMB67EH was used for VSV-G-tagging and expression of ACIAD2681 (ref. 26).

Protein secretion. Protein supernatants were analysed as described previously³. Immunoblotting followed standard methodology using anti-VSV-G primary antibodies (Sigma). The recombinantly expressed ACIAD2681 PAAR protein carried a standard VSV-G epitope tag at its C terminus.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 March; accepted 12 July 2013.

Published online 7 August 2013.

- Pukatzki, S. et al. Identification of a conserved bacterial protein secretion system in *Vibrio cholerae* using the *Dictyostelium* host model system. *Proc. Natl Acad. Sci. USA* **103**, 1528–1533 (2006).
- Mougous, J. D. et al. A virulence locus of *Pseudomonas aeruginosa* encodes a protein secretion apparatus. *Science* **312**, 1526–1530 (2006).
- Basler, M., Pilhofer, M., Henderson, G. P., Jensen, G. J. & Mekalanos, J. J. Type VI secretion requires a dynamic contractile phage tail-like structure. *Nature* **483**, 182–186 (2012).
- Kapitein, N. & Mogk, A. Deadly syringes: type VI secretion system activities in pathogenicity and interbacterial competition. *Curr. Opin. Microbiol.* **16**, 52–58 (2013).
- Basler, M., Ho, B. T. & Mekalanos, J. J. Tit-for-tat: type VI secretion system counterattack during bacterial cell-cell interactions. *Cell* **152**, 884–894 (2013).
- Pukatzki, S., Ma, A. T., Revel, A. T., Sturtevant, D. & Mekalanos, J. J. Type VI secretion system translocates a phage tail spike-like protein into target cells where it cross-links actin. *Proc. Natl Acad. Sci. USA* **104**, 15508–15513 (2007).

- Leiman, P. G. et al. Type VI secretion apparatus and phage tail-associated protein complexes share a common evolutionary origin. *Proc. Natl Acad. Sci. USA* **106**, 4154–4159 (2009).
- Dong, T. G., Ho, B. T., Yoder-Himes, D. R. & Mekalanos, J. J. Identification of T6SS-dependent effector and immunity proteins by Tn-seq in *Vibrio cholerae*. *Proc. Natl Acad. Sci. USA* **110**, 2623–2628 (2013).
- Kostyuchenko, V. A. et al. Three-dimensional structure of bacteriophage T4 baseplate. *Nature Struct. Biol.* **10**, 688–693 (2003).
- Kanamaru, S. et al. Structure of the cell-puncturing device of bacteriophage T4. *Nature* **415**, 553–557 (2002).
- Rossmann, M. G. & Blow, D. M. The detection of sub-units within the crystallographic asymmetric unit. *Acta Crystallogr.* **15**, 24–31 (1962).
- Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
- Kikuchi, Y. & King, J. Genetic control of bacteriophage T4 baseplate morphogenesis. II. Mutants unable to form the central part of the baseplate. *J. Mol. Biol.* **99**, 673–680 (1975).
- Marchler-Bauer, A. et al. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**, D348–D352 (2013).
- Hamburger, Z. A., Brown, M. S., Isberg, R. R. & Bjorkman, P. J. Crystal structure of invasin: a bacterial integrin-binding protein. *Science* **286**, 291–295 (1999).
- Felisberto-Rodrigues, C. et al. Towards a structural comprehension of bacterial type VI secretion systems: characterization of the TssJ-TssM complex of an *Escherichia coli* pathovar. *PLoS Pathog.* **7**, e1002386 (2011).
- Koskineni, S. et al. Rhs proteins from diverse bacteria mediate intercellular competition. *Proc. Natl Acad. Sci. USA* **110**, 7032–7037 (2013).
- Hachani, A. et al. Type VI secretion system in *Pseudomonas aeruginosa*: secretion and multimerization of VgrG proteins. *J. Biol. Chem.* **286**, 12317–12327 (2011).
- McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
- Adams, P. D. et al. The Phenix software for automated determination of macromolecular structures. *Methods* **55**, 94–106 (2011).
- Murshudov, G. N. et al. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Crystallogr. D* **67**, 355–367 (2011).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Finn, R. D., Clements, J. & Eddy, S. R. HMMER web server: interactive sequence similarity searching. *Nucleic Acids Res.* **39**, W29–W37 (2011).
- Söding, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res.* **33**, W244–W248 (2005).
- Geer, L. Y., Domrachev, M., Lipman, D. J. & Bryant, S. H. CDART: protein homology by domain architecture. *Genome Res.* **12**, 1619–1623 (2002).
- Fürste, J. P. et al. Molecular-cloning of the plasmid RP4 primase region in a multi-host-range *TacP* expression vector. *Gene* **48**, 119–131 (1986).

Acknowledgements We thank the entire staff of the Swiss Light Source PX beamlines and V. Olieric in particular for support related to crystallographic data collection. The project was made possible by the Swiss National Science Foundation (grant 31003A_127092) and EPFL funding to P.G.L., and grants AI-026289 and AI-01845 from the NIAID to J.J.M.

Author Contributions M.M.S. performed the initial bioinformatic analysis that led to the identification of PAAR proteins; M.M.S. cloned all VgrG and PAAR proteins for biochemical characterization and crystallization; M.M.S. and P.G.L. designed VgrG-PAAR binding experiments; M.M.S. and S.A.B. performed gp5(VgrG)-PAAR binding experiments and purification of gp5(VgrG)-PAAR complexes; S.A.B. crystallized gp5(VgrG)-PAAR complexes and determined their crystal structures; J.J.M., M.B. and B.T.H. designed T6SS secretion and T6SS-mediated killing assay experiments involving PAAR mutants; B.T.H. and M.B. performed these experiments; B.T.H. performed the bioinformatics analysis of PAAR protein domain extensions. All authors participated in writing the manuscript.

Author Information The atomic coordinates and the structure factors of the refined atomic models of gp5(VCA0018)-VCA0105 and gp5(c1883)-c1882 complexes were deposited to the Protein Data Bank (<http://www.rcsb.org>) under the accession numbers 4JIV and 4JIW, respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.G.L. (petr.leiman@epfl.ch) or J.J.M. (john_mekalanos@hms.harvard.edu).

METHODS

Cloning and protein expression. The gp5 fragment containing residues 484–575 was cloned into the in-house designed expression vector pEEva2 with using the In-Fusion system (Clontech). The pEEva2 vector is a derivative of the pET-23a expression plasmid (Novagen) with a modified multiple cloning region. The T7-tag was replaced with a MGSSH₆SSG polyhistidine tag (His-tag) followed by a sequence containing a tobacco etch virus (TEV) protease cleavage site (ENLYFQ⁸GSGS, cleavage site denoted by ⁸). Upon expression and after TEV cleavage, the gp5 fragment contained four residues (GSGS) upstream of residue 484. The C-terminal residues of gp5 were modified by performing PCR with long primers containing the required mutations (Extended Data Fig. 1).

For expression in *cis*, PAAR genes were cloned downstream from the gp5 fragment. The spacer between the stop codon of gene 5 and the first codon of PAAR genes contained a standard ribosome binding sequence and was about 15 bases long.

For expression in *trans*, PAAR genes were cloned into the pATE vector (chloramphenicol selection), which is a modified version of the pACYCDuet-1 expression plasmid (Novagen). The dual cloning site of the pACYCDuet-1 vector was replaced by the multiple cloning site from the pEEva2 vector. The cloned PAAR proteins retained their wild type N and C termini (that is, were tag-free).

Expression and purification of gp5(VgrG)-PAAR complexes. gp5(VgrG)-PAAR complexes were expressed in *Escherichia coli* B834(DE3) cells grown in the 2×TY medium containing ampicillin at 100 µg ml⁻¹ for the *cis* constructs and, additionally, chloramphenicol at 34 µg ml⁻¹ for the *trans* constructs. The cultures with a total volume of 2 litres were incubated at 37 °C with shaking at 200 r.p.m. until the culture reached $D_{600\text{ nm}}$ of 0.6. The culture was cooled down to 18 °C, and the protein expression was induced by an addition of isopropyl-β-D-thiogalactoside to a final concentration of 1 mM. The expression continued overnight.

The cells were collected on the following day by centrifugation at 8,000g at 4 °C for 10 min. The cell pellet was resuspended in a lysis buffer that contained 50 mM TrisCl pH 8.0, 300 mM NaCl, 5 mM imidazole. The cells were lysed by ultrasonication, which was performed on ice with the temperature of the lysate maintained below 10 °C. The lysate was then centrifuged at 25,000g, at 4 °C for 10 min. The supernatant was applied to a Ni column (5 ml GE HisTrap FF Crude). The non-specifically bound material was removed by washing the column with 10 column volumes of a washing buffer (50 mM TrisCl pH 8.0, 300 mM NaCl, 20 mM imidazole). The affinity bound material was eluted with 10 column volumes of an elution buffer (20 mM TrisCl pH 8.0, 300 mM NaCl, 250 mM imidazole).

The fractions containing the target protein were pulled together and the TEV protease was added to have a concentration of 10% (w/w) of the target protein. The mixture was then placed into a dialysis bag, and the proteolysis continued overnight with a simultaneous dialysis against a buffer optimal for TEV cleavage (10 mM TrisCl pH 8.0, 3 mM DTT (dithiothreitol), 1.5 mM EDTA). The digested protein was further purified by ion-exchange chromatography performed with a GE Mono Q 10/100 GL column connected to an AKTApurifier 100 system (GE Healthcare Life Sciences). The sample was loaded onto the column that was equilibrated with buffer A (20 mM TrisCl pH 8.0) and eluted with a linear gradient against buffer B (20 mM TrisCl pH 8.0, 1 M NaCl). The gradient extended from 0% to 65% of the buffer B concentration. Relevant fraction were combined and concentrated using Sartorius ultrafiltration devices with a molecular weight cutoff of 10,000 to a volume of ~5 ml. This sample was then loaded onto a GE HiLoad 16/60 Superdex 200 size-exclusion column pre-equilibrated with 10 mM TrisCl pH 8.0, 150 mM NaCl. The fractions containing pure gp5(VgrG)-PAAR complexes were combined and concentrated to 25 mg ml⁻¹ with the help of a similar Sartorius ultrafiltration devices and without changing the buffer. The protein was stored in the same buffer at 4 °C until it was used for crystallization. All purification buffers and the final protein solution contained NaN₃ at a concentration of 0.02% (w/v).

Identification of gp5(VgrG)-PAAR complexes. The purification procedure employed metal affinity chromatography based on the N-terminal His-tags of gp5 fragments with slightly modified C termini (Extended Data Fig. 1), and thus led to accumulation of two protein species, gp5(VgrG)-PAAR complexes and gp5 fragments alone. The β-helical fragment of gp5 used in this study does not fully denature in SDS-PAGE, and runs as several species with molecular mass of 10–30 kDa. Most of the PAAR proteins (molecular mass ~9.7 kDa) had the same SDS-PAGE mobility as the fastest migrating band of gp5 fragments (molecular mass ~10 kDa). However, these species were easily separable by a combination of high resolution anion exchange and analytical size exclusion chromatographies (monoQ and Superdex 75 resins, respectively). The gp5 fragment is a highly negatively charged β-helix¹⁰ that binds to the monoQ anion exchange resin strongly. Consequently, gp5 fragments that were free from PAAR proteins eluted from the monoQ resin by a buffer with a specific conductivity of 24–29 mS cm⁻¹. All gp5(VgrG)-PAAR complexes discussed in this paper were found to be much weaker anionic binders and eluted from the monoQ resin by a buffer with a

specific conductivity of 18–25 mS cm⁻¹. Furthermore, size exclusion profiles (Superdex 75 10/300 GL) of all gp5(VgrG)-PAAR complexes had only one peak centred at 10.34 ± 0.06 ml, whereas gp5 fragments free from PAAR proteins formed various oligomers and eluted at different positions of the chromatogram. The VCA0284 PAAR protein containing the transthyretin domain was prone to aggregation and the gp5(VCA0018)-VCA0284 complex was unsuitable for size-exclusion chromatography. Expression of VCA0284 in the presence of the gp5(VCA0018)-VCA0018 or gp5(VCA0018)-VCA0123 chimaeras led to an almost complete loss of gp5 from solution, indicating complex formation.

Crystallization and structure determination of gp5(VgrG)-PAAR complexes. The initial crystallization screening was carried out by the sitting drop method in 96-well SWISSCI 2-lens MRC plates using Jena Bioscience crystallization screens. Bio-Tek Precision XS and TTP LabTech mosquito pipetting robots were employed for preparing crystallization plates and setting up drops each containing 200 nl of the protein and the same volume of the well solution. Optimization of crystallization conditions was performed in 24-well SuperClear plates and thick siliconized cover slides (both from Jena Bioscience) by hanging drop vapour diffusion. Crystallization drops of the 24-well plate setup contained 1.25 µl of the protein solution in 10 mM TrisCl pH 8.0, 150 mM NaCl mixed with an equal volume of the well solution. Best crystals of the gp5(c1883)-c1882 complex were obtained with the protein having the initial concentration of 13.5 mg ml⁻¹ and equilibrated against 500 µl of the well solution containing 100–150 mM CaCl₂, 13–15% PEG 3350, 100 mM MES pH 6.5. Best crystals of the gp5(VCA0018)-VCA0105 complex were obtained with the protein at 15 mg ml⁻¹ and equilibrated against 500 µl of the well solution containing 13–14% PEG 2000, 100 mM NaAc pH 5.0.

For data collection, the crystals were dipped for 20–45 s into cryo solutions containing either 30% of glycerol for gp5(c1883)-c1882 or 25% of 2-methyl-2,4-pentanediol for gp5(VCA0018)-VCA0105 in addition to the well solution components and flash frozen in a vaporized nitrogen stream at 100 K. Collection of diffraction data and fluorescent scans was carried out at the PXI and PXIII beam lines of the Swiss Light Source (SLS) at the Paul Scherrer Institute (Villigen, Switzerland), respectively, using X-rays with a wavelength of 1.000 Å for both crystals. Best gp5(c1883)-c1882 and gp5(VCA0018)-VCA0105 crystals diffracted to 3.4 Å and 1.9 Å resolution limits, respectively. The diffraction data was indexed, integrated and scaled with XDS²⁷ (Extended Data Table 1). The structure of the gp5(c1883)-c1882 complex was solved by molecular replacement with PHASER¹⁹ using a fragment of the gp5-gp27 complex crystal structure containing residues 484–575 of gp5 (PDB ID 1K28) as a search model¹⁰. As the asymmetric unit contained four gp5(c1883)-c1882 complexes, non-crystallographic symmetry (NCS) averaging was used to improve the electron density before model building. The model was built manually with COOT²⁸ and refined with REFMAC5 (ref. 21) and PHENIX²⁰ using NCS for torsion angles restraints. The structure of gp5(c1883)-c1882 was subsequently used as a search model to solve the structure of the gp5(VCA0018)-VCA0105 complex by molecular replacement with PHASER¹⁹. There was only one complex per asymmetric units. The structure was refined with PHENIX²⁰ and COOT^{22,28}.

Both refined structures, gp5(VCA0018)-VCA0105 and gp5(c1883)-c1882, had excellent stereochemical parameters (see also Extended Data Table 1). All main chain dihedral angles of the gp5(VCA0018)-VCA0105 structure were found in the most favoured regions of the Ramachandran plot. The gp5(c1883)-c1882 structure had 97.3% of its main chain dihedral angles in the most favoured regions of the Ramachandran plot, 2.7% in the additionally allowed regions, and none in the disallowed region.

Bioinformatics. Putative orthologues of PAAR repeat proteins were identified with the help of HMMER²³. A search for high confidence hits to PFAM²⁹ domains PF05488 or PF13665 that describe the two distinct PAAR repeat domains resulted in 1353 PAAR domain-containing proteins. Phage proteins and orphan genes were excluded from this analysis. Functional assignment of extension domains was based on high confidence hits obtained with the help HHpred²⁴, HMMER²³, CDD¹⁴ and CDART²⁵ web services.

Bacterial killing assays. Overnight cultures of *V. cholerae* 2740-80, *A. baylyi* ADP1 and *E. coli* MG1655 were washed in LB and diluted 50–200× into fresh LB and cultivated for 2.5–3.5 h to reach $D_{600\text{ nm}}$ of ~0.5–1.0. Cells were pelleted by centrifugation at 4,000g for 5 min, mixed at $D_{600\text{ nm}}$ ~10 in 1:1 ratio, and 5 µl of the mixture was spotted on a dry LB agar plate. After 2 h at 37 °C, bacterial spots were cut out and the cells were resuspended in 1 ml LB. The cellular suspension was serially diluted in LB, and 5 µl of the suspensions was spotted on selective plates (streptomycin for *V. cholerae* and *A. baylyi*, and gentamicin for *E. coli*). *aciad2681* was expressed basally from the plasmid pMMB67EH²⁶. PAAR protein complementation in *V. cholerae* was done by cloning the indicated PAAR genes into pBAD24 (ref. 30) and inducing with the appropriate concentrations of L-arabinose. Two-tailed Student's *t*-test was used to interpret the statistical significance of the number of the colony forming units.

Protein secretion. Overnight cultures of *V. cholerae* 2740-80 and *A. baylyi* ADP1 were washed in LB and diluted 200× into fresh LB and cultivated for 5 h at 37 °C. One ml of culture supernatants were passed through a 0.2-μM filter, precipitated with TCA, subjected to 12% SDS-PAGE and stained with Coomassie blue. The recombinantly expressed ACIAD2681 PAAR protein carried a C-terminal VSV-G epitope tag (YTDIEMNRLGK). Western blots were performed on the TCA precipitated samples following standard methodology using anti-VSV-G primary antibodies (Sigma). Band intensities were quantified using Fiji software³¹.

Molecular graphics. UCSF Chimera³² was used to prepare Fig. 1 and Extended Data Figs 2–6.

27. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).

28. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).

29. Finn, R. D. *et al.* The Pfam protein families database. *Nucleic Acids Res.* **38**, D211–D222 (2010).

30. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–4130 (1995).

31. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nature Methods* **9**, 676–682 (2012).

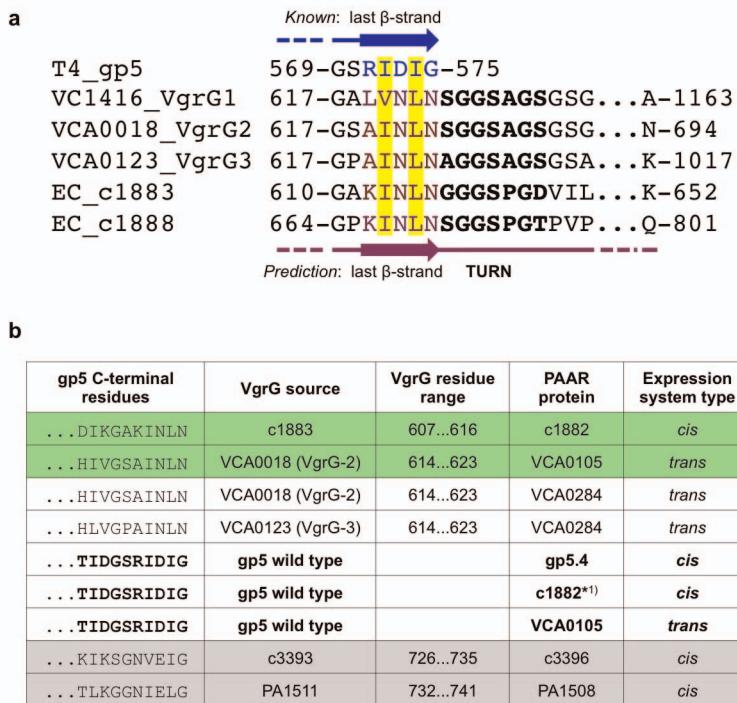
32. Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).

33. Kyte, J. & Doolittle, R. F. A simple method for displaying the hydrophobic character of a protein. *J. Mol. Biol.* **157**, 105–132 (1982).

34. Bearden, J. A. X-ray wavelengths. *Reviews of Modern Physics* **39**, 78–124 (1967).

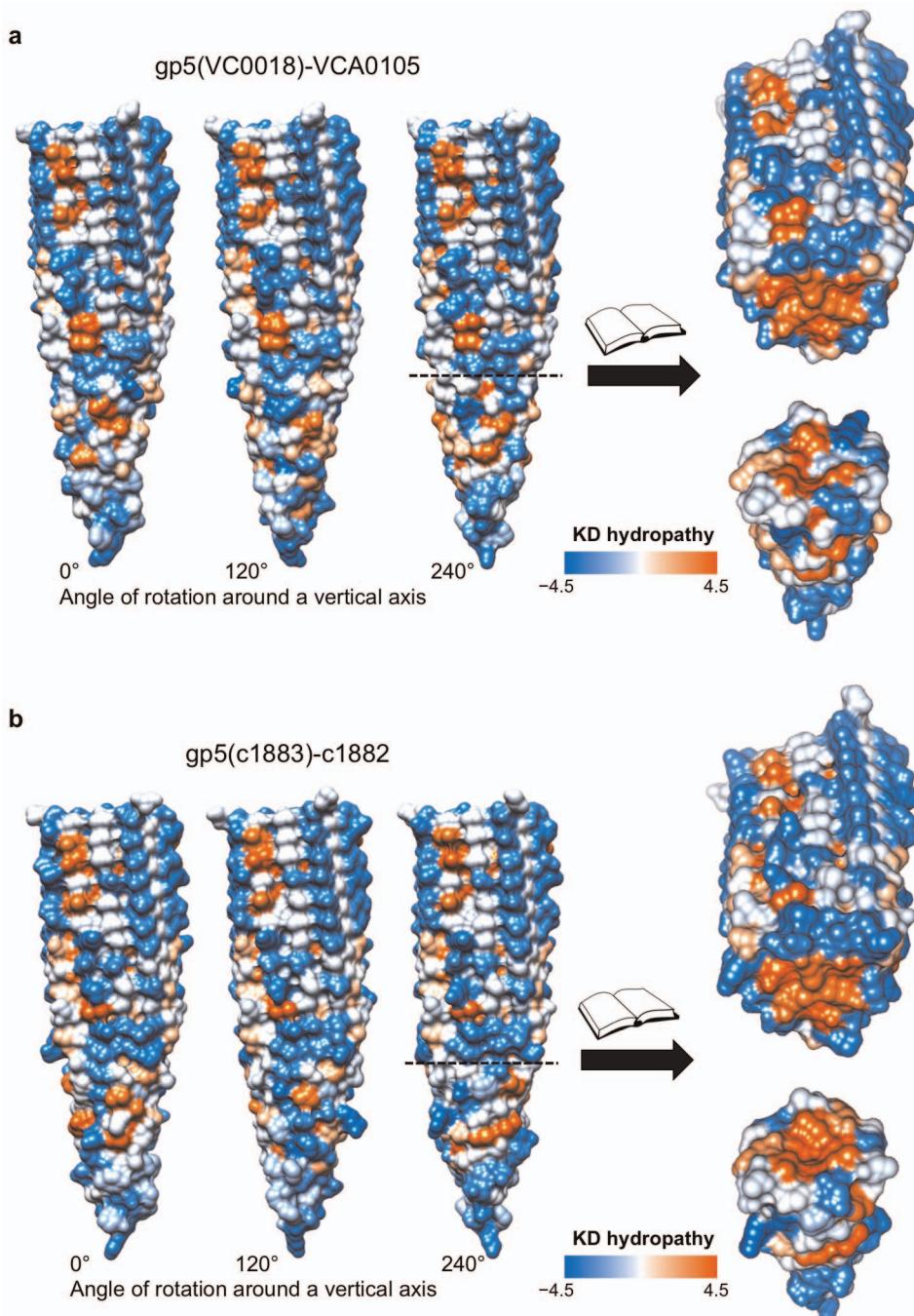
35. Crooks, G. E., Hon, G., Chandonia, J. M. & Brenner, S. E. WebLogo: a sequence logo generator. *Genome Res.* **14**, 1188–1190 (2004).

36. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).



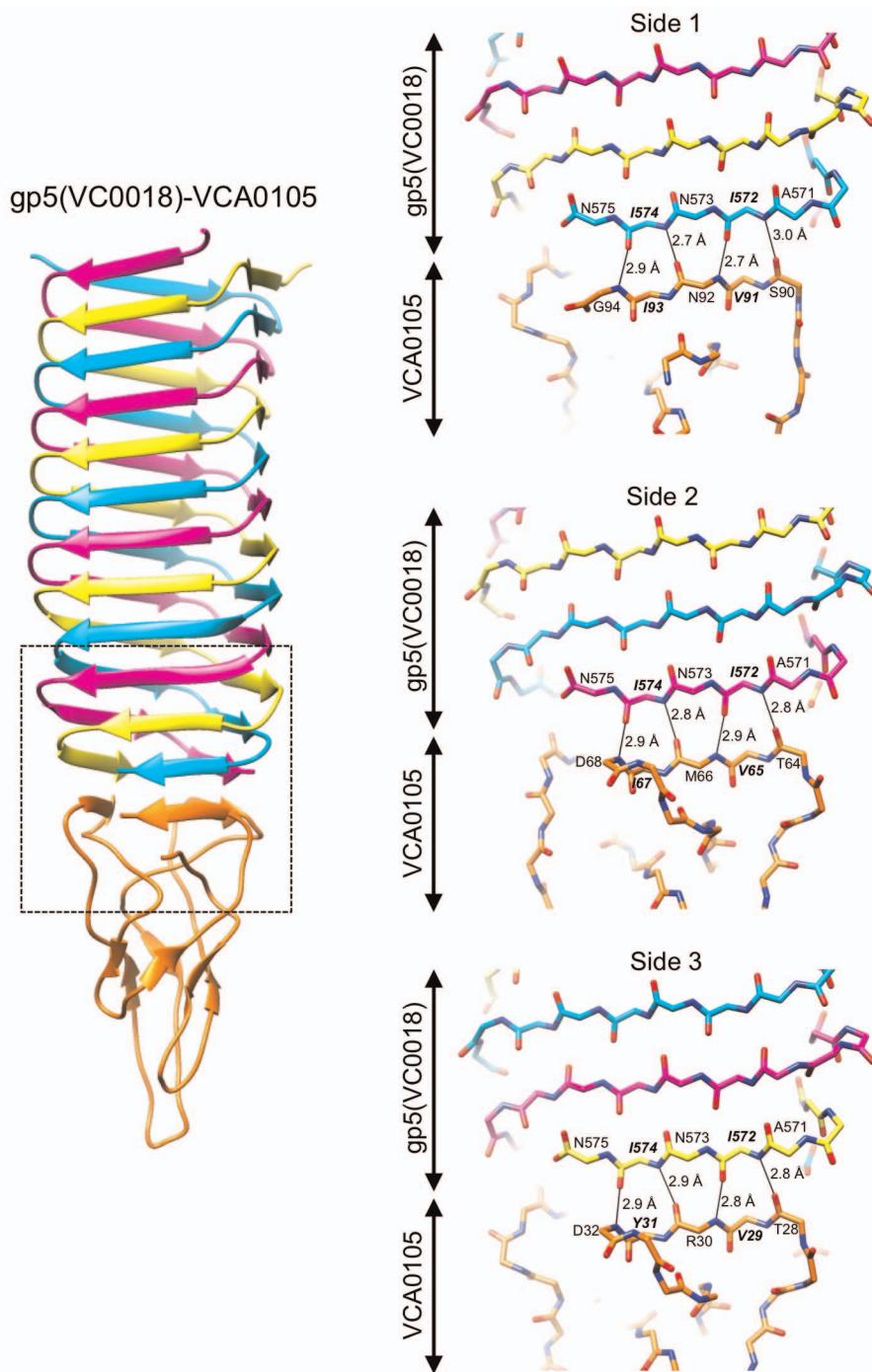
Extended Data Figure 1 | Design of gp5–VgrG chimaeras and analysis of their interaction with selected PAAR proteins. **a**, Identification of the blunt ends of selected VgrG β -helices. The last β -strand of the known T4 gp5 structure is shown in blue. Putative β -strands terminating VgrG β -helices are in plum colour. The glycine/serine-rich motif is bold highlighted. The residue number for the first amino acid of the shown fragment and that for the C-terminal amino acid of the protein are given. Abbreviations used EC, *E. coli* CTF073; VC, *V. cholerae* O1 biovar El Tor str. N16961. **b**, Binding of several

PAAR proteins to gp5-based VgrG-like β -helices. Two complexes for which the crystal structures are reported in this paper are highlighted with green background. Entries showing gp5 modifications that did not result in PAAR binding have a grey background. c1882* contains three mutations T28K, T64K and T90K that were made to mimic T4 gp5.4. Note that VCA0105 and VCA0284 contain threonines and serines in these positions and they bind wild-type gp5.



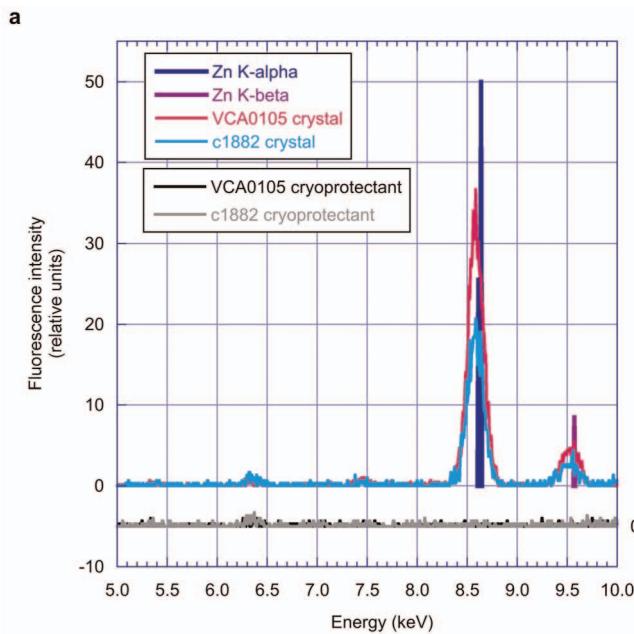
Extended Data Figure 2 | Surface features of gp5(VgrG)-PAAR complexes and VgrG-PAAR interface. **a, b,** Molecular surfaces are coloured according to their hydrophobicity with sky blue, white, and orange corresponding to the most hydrophilic, neutral and hydrophobic patches, respectively. Residue

hydrophobicity values are according to the Kyte-Doolittle scale³³, which is given as a coloured bar labelled 'KD hydrophathy'. Three orientations of the gp5(VgrG)-PAAR complex and an 'open book' view of the VgrG-PAAR interface are shown for both PAAR proteins.



Extended Data Figure 3 | Main chain hydrogen-bonding network of VgrG-PAAR interface. The dashed line rectangle in the left panel indicates the area shown enlarged in the panels on the right. The three right panels show the main chain hydrogen bonds between VgrG and PAAR proteins for three different

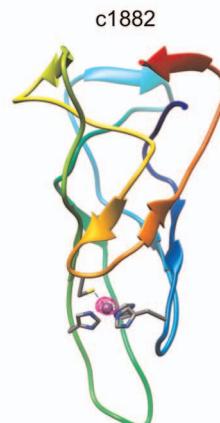
sides of the gp5(VgrG)-PAAR complex. Residues with side chains pointing inwards and forming the VgrG-PAAR hydrophobic interface are in bold italic. Side chains are not shown for clarity. The carbon atoms of the PAAR protein are colored orange.



b

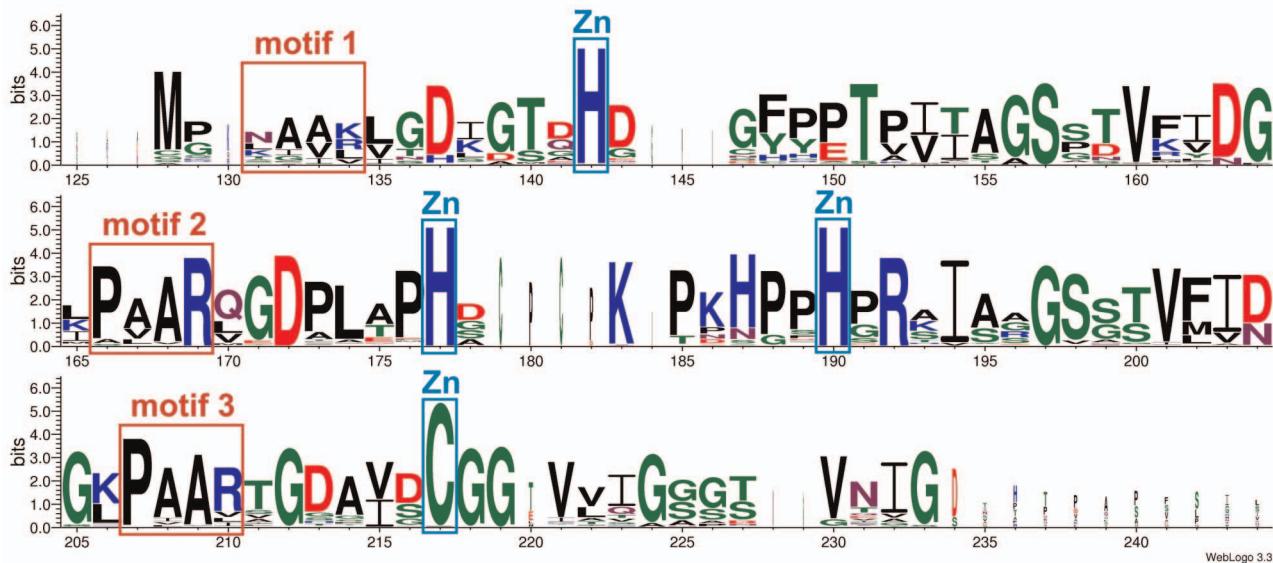


c

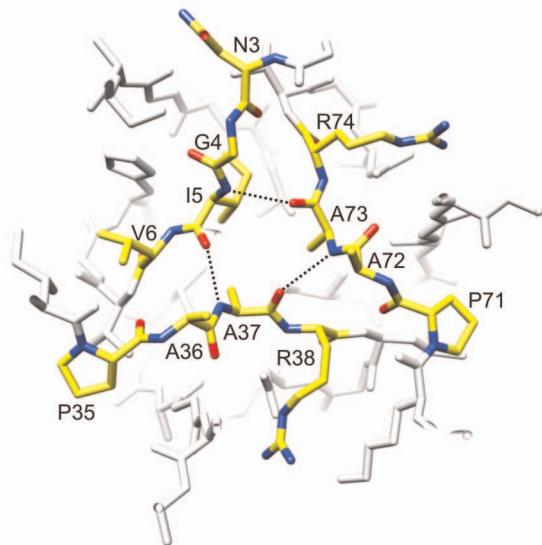


Extended Data Figure 4 | PAAR proteins contain a Zn atom. **a**, X-ray fluorescence spectra of gp5(VC0018)-VCA0105 and gp5(c1883)-c1882 crystals and their cryoprotectant solutions. The excitation wavelength is 1.0 Å (~ 12.4 keV). The energies and peak heights for the atomic Zn $K\alpha$ and $K\beta$ lines are taken from ref. 34. **b, c**, Anomalous difference Fourier maps (magenta

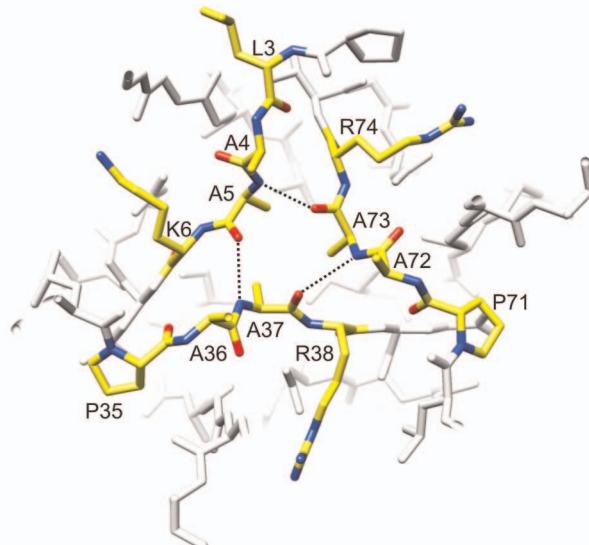
mesh) of both crystal structures contain only one non-noise peak corresponding to the Zn atom. The VCA0105 and c1882 maps are contoured at 15.0 and 6.0 standard deviations above the mean, respectively. The corresponding noise level of the two maps is ~ 5.1 and ~ 4.5 standard deviations above the mean, respectively.

a**b**

VCA0105

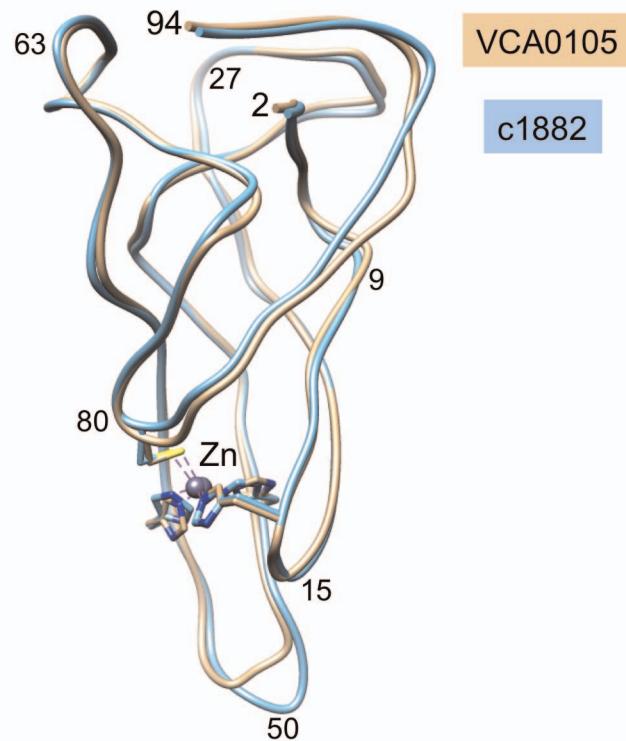
**c**

c1882

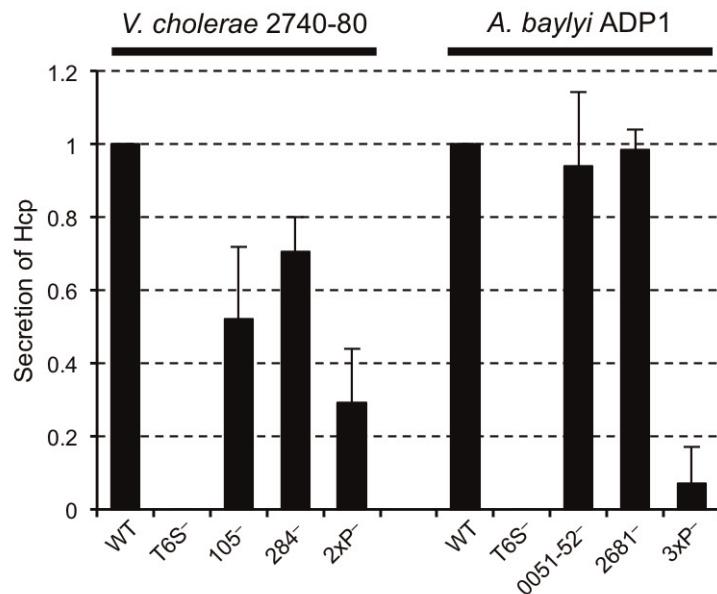
**Extended Data Figure 5 | Conserved features of PAAR proteins.**

a, WebLogo³⁵ sequence alignment of VCA0105 homologues identified with BLAST³⁶. GenBank accession numbers of protein sequences that were used to prepare this WebLogo diagram in order of their appearance in the BLAST alignment file: 15600876, 227811731, 153212840, 417818856, 262403304, 258622274, 261212686, 260769543, 336125205, 417954065, 153801242, 424808850, 399908303, 15601050, 153817580, 254284956, 444380739, 307545498, 237731059, 197336227, 387815895, 59712519, 149377963, 120556511, 359395272, 338997855, 209695798, 338998698, 126665853, 440287404, 327412950, 145298557, 385332665, 87118978, 433087003, 433201317, 433072538, 432504178, 222155056, 262169027, 26247740, 145300690, 422833042, 432730965, 91210698, 432592619, 222156204, 432758040, 218704924, 26247745, 222156197, 417084314, 50121038,

444375671, 253990116, 423141622, 416895592, 423207453, 417628384, 331657463, 406676759, 251790158, 323491128, 421082032, 437829594, 120555573, 433076042, 145301291, 269102718, 261823483, 343512222, 343512225, 343515540, 343510421, 416895589, 145299403, 117619593, 385873501, 425070534, 343512226, 343510889, 422021929, 229523366, 262402814, 385870158, 411009627, 385869971, 258627229, 295418810, 37525853, 261819862, 307133118, 432995158, 422803770, 253987746, 253688569, 114319223, 268590359 and 261822686. The conserved PAAR motif and residues forming the Zn binding site are labelled. **b, c**, Pseudotrimeric organization of the three interacting PAAR motifs in VCA0105 and c1882, respectively. The buried hydrogen bonds stabilizing the fold are shown with dashed lines.

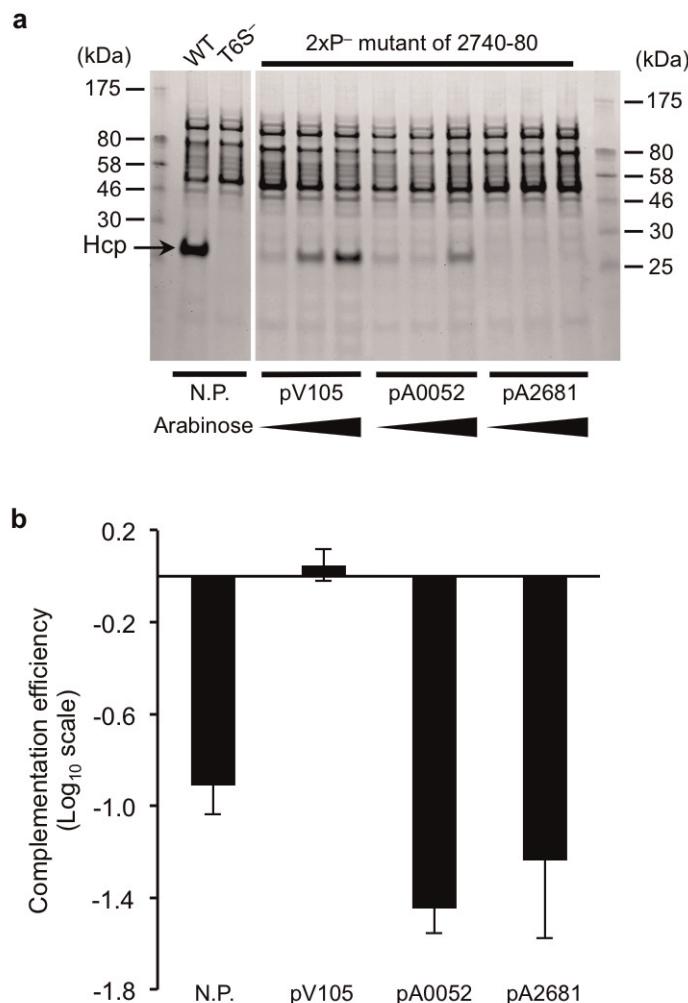


Extended Data Figure 6 | Superposition of VCA0105 and c1882 PAAR structures. Residue numbers are given at strategic positions.



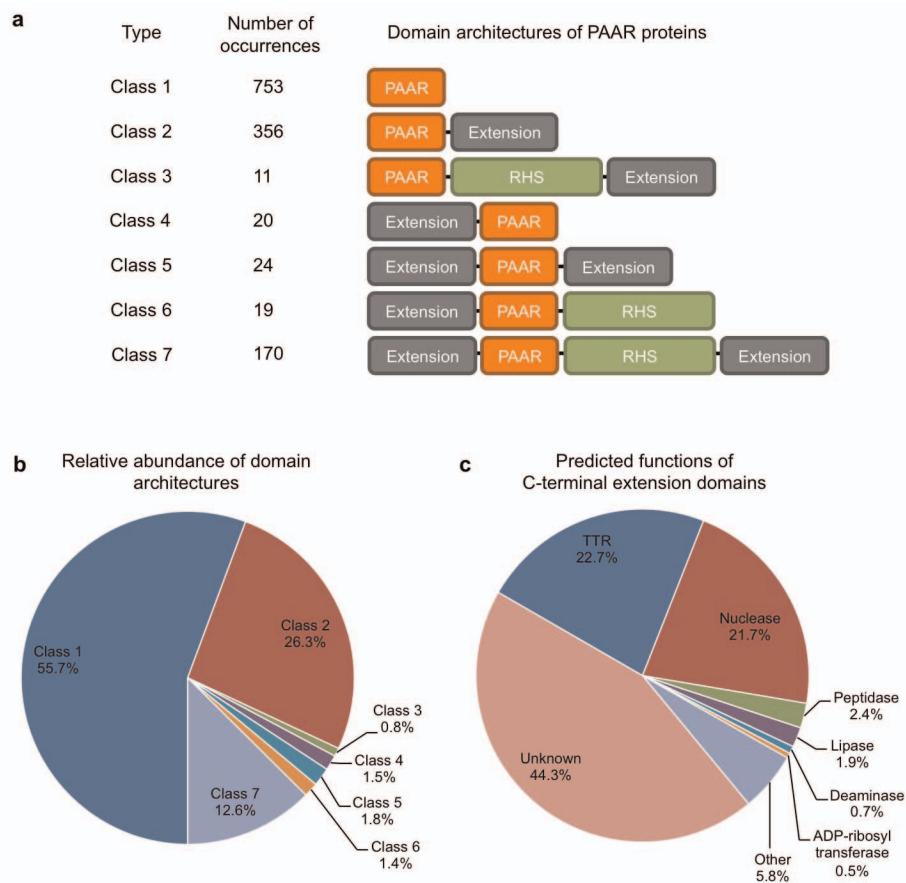
Extended Data Figure 7 | Quantification of Hcp secretion. Hcp secretion was evaluated by quantifying the intensity of the Hcp band in SDS-PAGE using the Fiji software³¹ and normalizing it to the wild type. The presented data are

the average of three experiments. The error bars correspond to one standard deviation. The mutants are labelled as in Fig. 2.



Extended Data Figure 8 | Complementation analysis of heterologous PAAR proteins. **a**, SDS-PAGE assay for Hcp secretion of the *V. cholerae* 2740-80 $2 \times P^-$ mutant described in Fig. 2 legend carrying a plasmid expressing endogenous or heterologous PAAR proteins. Plasmids pV105, pA0052, and pA2681 express PAAR genes *vca0105* (*Vibrio*), *aciad0052* (*Acinetobacter*) and *aciad2681* (*Acinetobacter*), respectively, under the control of arabinose-inducible pBAD promoter³⁰. N.P. stands for no plasmid. Cells were grown in the presence of 0.001%, 0.01% and 0.1% arabinose for 4 h. Both panels show a

typical representative of three identical experiments. **b**, *E. coli* MG1655 killing by the *V. cholerae* 2740-80 $2 \times P^-$ mutant that carries a plasmid expressing endogenous and heterologous PAAR proteins. The complementation efficiency is calculated as the \log_{10} ratio of *E. coli* recovered after competition with the wild type strain of *V. cholerae* 2740-80 and its $2 \times P^-$ mutant supplemented with the indicated plasmid. The data represent the average of three experiments; error bars correspond to one standard deviation.



Extended Data Figure 9 | Bioinformatic analysis of PAAR proteins. **a**, Domain organization of all known bacterial (non-phage) PAAR proteins. **b**, Relative abundance of the seven distinct domain organizations. **c**, Predicted functions of the C-terminal domains. TTR stands for transthyretin domain.

Extended Data Table 1 | Crystallographic data collection and refinement statistics.

	gp5(VCA0018)–VCA0105	gp5(c1883)–c1882
Data collection		
Space group	P4 ₂ 22	P6 ₁ 22
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	113.7, 113.7, 76.9	187.16, 187.16, 238.89
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 120.0
Resolution (Å)	80.4 – 1.9 (2.0 – 1.9)*	48.6 – 3.4 (3.6 – 3.4)*
<i>R</i> _{meas}	0.06 (0.37)	0.15 (1.01)
<i>I</i> / σ <i>I</i>	20.0 (4.9)	12.6 (2.1)
Completeness (%)	99.9 (99.3)	99.8 (98.9)
Redundancy	6.9 (7.0)	7.0 (6.7)
Refinement		
Resolution (Å)	80.4 – 1.9	48.6 – 3.4
No. reflections	40054	33366
<i>R</i> _{work} / <i>R</i> _{free}	0.16 / 0.21	0.18 / 0.24
No. atoms		
Protein	2880	10848
Ligand/ion	60	11
Water	453	49
B-factors		
Protein	39.4	163.1
Ligand/ion	63.3	173.7
Water	49.3	122.4
R.m.s deviations		
Bond lengths (Å)	0.007	0.002
Bond angles (°)	0.995	1.039

Both data sets were collected using a single crystal. *Highest resolution shell is shown in parenthesis.

An Sp1 transcription factor coordinates caspase-dependent and -independent apoptotic pathways

Takashi Hirose¹ & H. Robert Horvitz¹

During animal development, the proper regulation of apoptosis requires the precise spatial and temporal execution of cell-death programs, which can include both caspase-dependent and caspase-independent pathways^{1,2}. Although the mechanisms of caspase-dependent and -independent cell killing have been examined extensively, how these pathways are coordinated within a single cell that is fated to die is unknown. Here we show that the *Caenorhabditis elegans* Sp1 transcription factor SPTF-3 specifies the programmed cell deaths of at least two cells—the sisters of the pharyngeal M4 motor neuron and the AQR sensory neuron—by transcriptionally activating both caspase-dependent and -independent apoptotic pathways. SPTF-3 directly drives the transcription of the gene *egl-1*, which encodes a BH3-only protein that promotes apoptosis through the activation of the CED-3 caspase³. In addition, SPTF-3 directly drives the transcription of the AMP-activated protein kinase-related gene *pig-1*, which encodes a protein kinase and functions in apoptosis of the M4 sister and AQR sister independently of the pathway that activates CED-3 (refs 4, 5). Thus, a single transcription factor controls two distinct cell-killing programs that act in parallel to drive apoptosis. Our findings reveal a bivalent regulatory node for caspase-dependent and -independent pathways in the regulation of cell-type-specific apoptosis. We propose that such nodes might act as features of a general mechanism for regulating cell-type-specific apoptosis and could be therapeutic targets for diseases involving the dysregulation of apoptosis through multiple cell-killing mechanisms.

The *C. elegans* pharyngeal M4 motor neuron is generated during embryonic development, whereas the M4 sister cell dies by apoptosis soon after its generation (Fig. 1a)^{6,7}. We constructed a *P_{ceh-28}::gfp* reporter transgene that expresses green fluorescent protein (GFP) specifically in the M4 neuron of wild-type animals and in both the M4 neuron and the surviving M4 sister of *ced-3* caspase mutants defective in apoptosis (Fig. 1b)⁸, allowing us to identify efficiently mutants defective in M4 sister cell death from a large-scale genetic screen. Among our isolates were two non-allelic mutations—*n4850* and *n4780*—which, on the basis of mapping and transformation-rescue studies, are alleles of the genes *sptf-3* and *pig-1*, respectively (Fig. 1b).

sptf-3 encodes an Sp1 family transcription factor with a characteristic glutamine-rich domain and three C2H2-type zinc finger domains (Fig. 1c). The *n4850* mutant has a single *sptf-3* mutation, at a splice acceptor site of the last exon (Fig. 1c). In total, 34% of *sptf-3(n4850)* mutants had a surviving M4 sister, and this cell-death defect was rescued by a transgene carrying only the *sptf-3* genomic locus (Fig. 1e). A deletion allele of *sptf-3*, *tm607Δ*, and inactivation of *sptf-3* by RNA interference (RNAi) both phenocopied the *sptf-3(n4850)* mutation, demonstrating that a reduction of *sptf-3* function causes a defect in M4 sister cell death (Fig. 1e).

pig-1 encodes an AMP-activated protein kinase (AMPK)-related protein kinase most similar to mammalian maternal embryonic leucine zipper kinase (MELK); *pig-1* is known to regulate the asymmetric cell divisions of several neuroblasts that divide to produce an apoptotic cell, including the M4 sister⁴ (Fig. 1d). The *n4780* mutant has a single *pig-1* mutation in the kinase domain, changing a conserved glycine at

amino acid 172 to glutamic acid (Fig. 1d). In total, 20% of *pig-1(n4780)* mutants had a surviving M4 sister, and this cell-death defect was rescued by a transgene carrying only the *pig-1* genomic locus (Fig. 1e). A presumptive null allele of *pig-1*, *gm344Δ*, and inactivation of *pig-1* by RNAi both phenocopied the *pig-1(n4780)* mutation, demonstrating that a reduction of *pig-1* function causes a defect in M4 sister cell death (Fig. 1e).

Both *sptf-3* and *pig-1* are required for the deaths of multiple cells, including the sisters of the AQR neuron, the pharyngeal gland cells 1A (g1A) and the pharyngeal I2 interneurons (Fig. 1f). By contrast, neither *sptf-3(n4850)* nor *pig-1(n4780)* affected the deaths of the sisters of the pharyngeal neurosecretory motor (NSM) neurons, the sisters of the pharyngeal I1 interneurons or the ventral C (VC) homologues of the ventral nerve cord (Fig. 1f). Thus, *sptf-3* and *pig-1* seem to promote apoptosis in the same subset of cells fated to die, suggesting that *sptf-3* and *pig-1* have a functional interaction in the regulation of cell death.

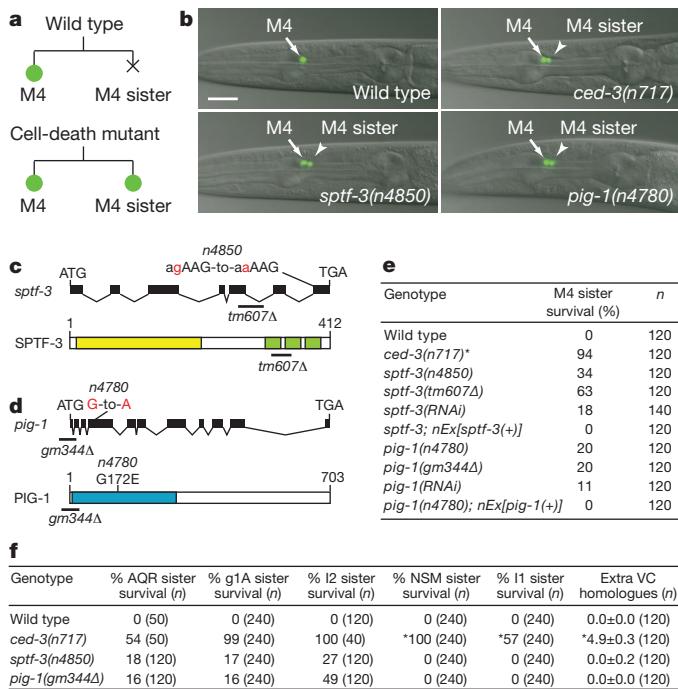


Figure 1 | *sptf-3* and *pig-1* promote the death of the M4 sister cell.

a, Schematic representation of the M4 cell lineage in the wild type and mutants defective in M4 sister cell death. b, Merged epifluorescence and Nomarski images of the pharynx in wild-type, *ced-3(n717)*, *sptf-3(n4850)* and *pig-1(n4780)* animals expressing *P_{ceh-28}::gfp*. Scale bar, 20 μm. c, d, Genomic organizations and protein structures of *sptf-3* (c) and *pig-1* (d), including the mutations *n4850* and *n4780*. The yellow and green boxes represent the Q-rich and C2H2-type zinc finger domains of SPTF-3, respectively. The blue box indicates the kinase domain of PIG-1. e, The percentages of M4 sister survival in animals of the indicated genotypes. f, The percentages of survival of the indicated cells and the number of extra VC homologues are shown. Errors are s.d. *Data from ref. 8.

¹Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, USA.

To identify direct transcriptional targets of SPTF-3 involved in the regulation of M4 sister cell death, we performed chromatin immunoprecipitation with massively parallel DNA sequencing (ChIP-seq) analyses using two different SPTF-3 polyclonal antibodies, N81 and M82, both of which specifically recognized the SPTF-3 protein (Supplementary Fig. 1). These experiments identified 2,459 genomic regions that immunoprecipitated with both antibodies (Supplementary Fig. 2a–d and Supplementary Tables 1, 2). Gene ontology analysis indicated that SPTF-3 functions in a variety of biological processes (Supplementary Fig. 2e), consistent with the observation that *sptf-3(n4850)*, the *sptf-3(tm607Δ)* deletion and *sptf-3* RNAi knockdown all cause cell-fate transformations, embryonic and larval lethality and morphological abnormalities (Supplementary Fig. 3 and Supplementary Fig. 4)⁹.

We identified an SPTF-3-bound region immediately upstream of the *pig-1* coding region (Fig. 2a). This region contains the consensus SPTF-3 binding motif (CGCCC) identified from our ChIP-seq analyses (Fig. 2b, c). We tested whether the SPTF-3 binding motif of the *pig-1* promoter region is necessary for *pig-1* to promote M4 sister cell death. A wild-type *pig-1* transgene rescued the M4 sister cell-death defect of *pig-1(n4780)* mutants, whereas neither a *pig-1* transgene lacking 71 base pairs of the SPTF-3-bound region of the *pig-1* promoter ($\Delta 71$ bp) nor a *pig-1* transgene containing mutations in the consensus SPTF-3 binding motif (mut.1) rescued the M4 sister cell-death defect of *pig-1(n4780)* mutants (Fig. 2d, e). The wild-type *pig-1*

promoter expressed GFP in many embryonic cells, whereas mutant *pig-1* promoters lacking the consensus SPTF-3 binding motif ($\Delta 71$ bp and mut.1) did not (Supplementary Fig. 5), indicating that the consensus SPTF-3 binding motif of the *pig-1* promoter region is required for *pig-1* expression. Furthermore, *pig-1* transcript levels in *sptf-3(n4850)* mutants were decreased by 43% compared to those of wild-type animals (Fig. 2f), and expression of a *P_{pig-1}::gfp* transgene was frequently absent from the seam cells, P cells, ALM neurons and AVM neuron of *sptf-3(n4850)* mutants (Fig. 2h and Supplementary Fig. 6). Conversely, overexpression of *sptf-3* from a multi-copy transgene under the control of the *sptf-3* promoter induced ectopic expression of *pig-1* in the seam cells and the hyp7 hypodermal cells (Fig. 2g, h). These results indicate that the consensus SPTF-3 binding motif of the *pig-1* promoter region is required for *pig-1* to promote M4 sister cell death and that SPTF-3 is necessary and sufficient for *pig-1* expression, suggesting that SPTF-3 directly drives *pig-1* expression in the regulation of M4 sister cell death.

Although SPTF-3 acts through *pig-1* to promote M4 sister cell death, our genetic observations suggested that *sptf-3* also functions through a pathway distinct from that of *pig-1*. A partial loss-of-function mutation of *sptf-3(n4850)*, caused a defect in M4 sister cell death more severe than that of the *pig-1* null mutation *gm344A* (Fig. 1e). The M4 sister cell-death defect of *sptf-3; pig-1* double mutants is much more severe than that of either single mutant (Supplementary Table 3). Therefore, we tested whether *sptf-3* acts through the pro-apoptotic BH3-only

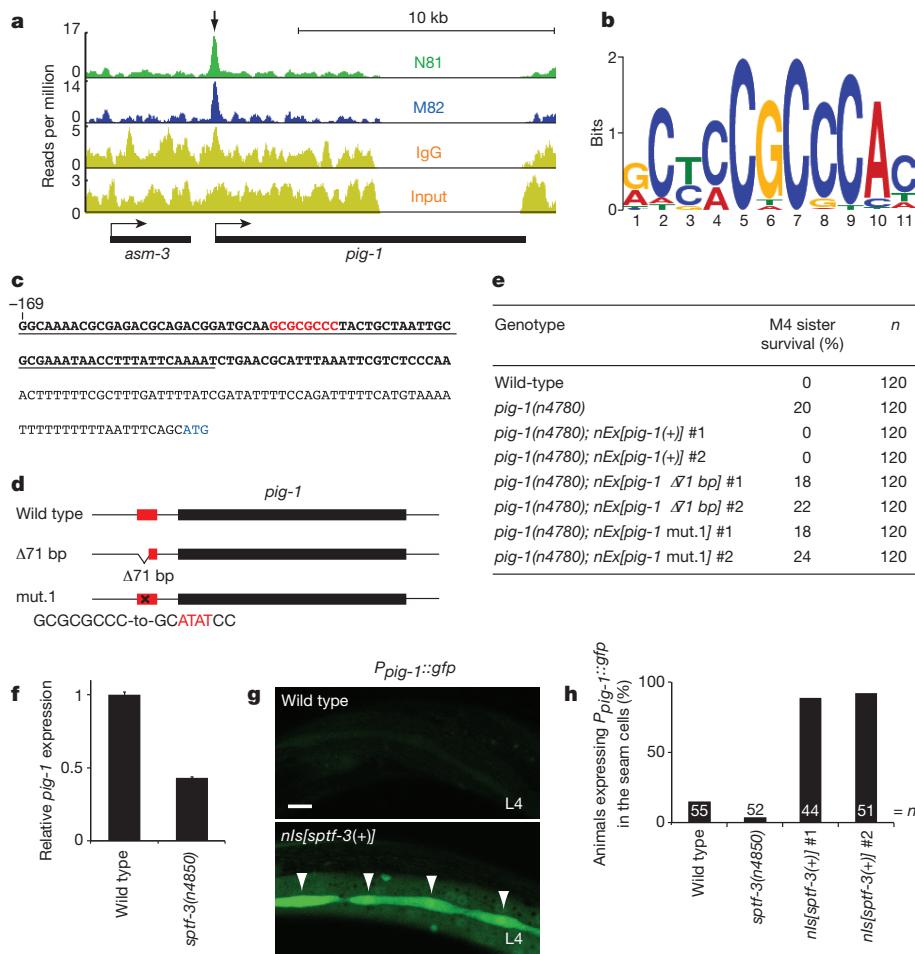


Figure 2 | *pig-1* is a direct transcriptional target of SPTF-3 in the regulation of the death of the M4 sister cell. **a**, ChIP-seq results obtained by immunoprecipitation with the antibodies anti-SPTF-3 N81, anti-SPTF-3 M82 or normal IgG and input chromatin are shown. Arrow, an enriched SPTF-3-bound region. **b**, The consensus SPTF-3 binding motif determined by ChIP-seq is represented. **c**, DNA sequence of the *pig-1* promoter. Bold letters, SPTF-3-bound regions identified by ChIP-seq. Red letters, the GC-rich sequence that contains the consensus SPTF-3 binding motif. Blue letters, the ATG start site of the *pig-1* coding region. Underline, the deleted *pig-1* sequences in the *pig-1(Δ71 bp)* transgene. **d**, Representations of the *pig-1* transgenes used in rescue experiments. Wild-type, a wild-type *pig-1* transgene; $\Delta 71$ bp lacks 71 base pairs of the SPTF-3-bound region; mut.1, mutated as shown for the consensus SPTF-3 binding motif. Red box, SPTF-3-bound regions identified by ChIP-seq. **e**, The percentages of M4 sister survival in animals of the indicated genotypes. **f**, *pig-1* messenger RNA levels of wild-type and *sptf-3(n4850)* embryos were measured by quantitative PCR with reverse transcriptase. *pig-1* expression levels of *sptf-3(n4850)* mutants relative to that of the wild type are represented. Error bars, s.d. of three technical replicates. **g**, Expression of *P_{pig-1}::gfp* in animals of the indicated genotypes at the fourth larval stage (L4). Arrowheads, seam cells. Scale bar, 20 μ m. **h**, The percentages of L4 larvae expressing *P_{pig-1}::gfp* in the seam cells of animals of the indicated genotypes.

gene *egl-1*, which functions in the caspase-dependent pathway of programmed cell death. *egl-1* acts through the anti-apoptotic *BCL2* homologue *ced-9*, the pro-apoptotic *APAF1* homologue *ced-4*, and the pro-apoptotic caspase gene *ced-3* to drive most cell deaths during the development of *C. elegans*³. As *egl-1* is required for M4 sister cell death⁸, we scored *egl-1* expression in the surviving M4 sister of *ced-3* mutants defective in apoptosis using a *P_{egl-1}::gfp* reporter transgene that expresses GFP under the control of the *egl-1* promoter⁸. *P_{egl-1}::gfp* was expressed in the M4 sister (100% of animals; *n* = 119) but not in the M4 neuron (0% of animals; *n* = 119) in *ced-3* mutants (Fig. 3a, b), indicating that *egl-1* is not only required for the apoptosis of, but also is expressed in, the M4 sister. We observed that only 33% of *sptf-3(tm607Δ)*; *ced-3* animals expressed *P_{egl-1}::gfp* in the M4 sister, whereas 100% of *pig-1(gm344Δ)*; *ced-3* animals expressed *P_{egl-1}::gfp* in the M4 sister (Fig. 3a, b). Thus, *sptf-3* but not *pig-1* is necessary for normal *egl-1* expression in the M4 sister, suggesting that *sptf-3* acts through both the *egl-1*-mediated apoptotic pathway and the *pig-1*-mediated apoptotic pathway, whereas *pig-1* acts through a pathway distinct from that of *egl-1* to promote M4 sister cell death.

We next asked whether SPTF-3 directly or indirectly drives *egl-1* expression in the M4 sister by examining the *egl-1* promoter region using our ChIP-seq data. We found a small but distinct SPTF-3 binding peak immediately upstream of an *egl-1* coding region (2.77- or 2.96-fold enrichment compared to an input control in the ChIP-seq experiments using SPTF-3 antibody N81 or M82, respectively) (Supplementary Fig. 7a). This SPTF-3-bound region contains two tandem consensus SPTF-3 binding motifs (GGGCAGGGCG) (Supplementary Fig. 7b). These results suggest that SPTF-3 binds to this region. As SPTF-3 regulates *egl-1* expression in a cell-type-specific manner (Fig. 3a and Supplementary Fig. 8) and whole embryos are used for ChIP-seq analyses, it is probable that SPTF-3 binds to the *egl-1* promoter region in only a small number of cells, resulting in a relatively small binding peak.

To test whether the SPTF-3-bound region of the *egl-1* promoter is important for *egl-1* to promote M4 sister cell death, we introduced wild-type and mutant *egl-1* transgenes into *egl-1(n1084 n3082)* mutants defective in M4 sister cell death. A wild-type *egl-1* transgene, pTH01, rescued the M4 sister cell-death defect of *egl-1* mutants, whereas an *egl-1* transgene lacking 30 bp of the SPTF-3-bound region of the *egl-1* promoter did not (Fig. 3c, d). To define further the *egl-1* promoter region important for M4 sister cell death, we introduced a series of mutations into this 30-bp region (Fig. 3c). An *egl-1* transgene containing mutations in the GC-rich sequence (mut.2), in the 9 bp next to the GC-rich

sequence at the 5' side (mut.1), in the 8 bp next to the GC-rich sequence at the 3' side (mut.3) or in both the GC-rich sequence and the 8 bp next to the GC-rich sequence at the 3' side (mut.4) all rescued the M4 sister cell-death defect of *egl-1* mutants. By contrast, an *egl-1* transgene containing mutations in both in the GC-rich sequence and 9 bp next to the GC-rich sequence at the 5' side (mut.5) did not rescue the M4 sister cell-death defect of *egl-1* mutants (Fig. 3d). These results indicate that an *egl-1* promoter sequence containing the consensus SPTF-3 binding motif is required for *egl-1* to promote M4 sister cell death, suggesting that SPTF-3 directly drives *egl-1* expression in the M4 sister.

To further test the hypothesis that SPTF-3 but not PIG-1 functions through *egl-1* in the regulation of M4 sister cell death, we performed epistasis analyses between *sptf-3* or *pig-1* and *ced-9*, which functions downstream of *egl-1*. Because the *ced-9(n2812)* null mutation causes ectopic cell deaths and organismal lethality, we used a weak *ced-3(n2446)* mutation in these experiments to suppress *ced-9(n2812)* lethality (Fig. 3e)¹⁰. *sptf-3*; *ced-9* double mutants had nearly the same penetrance of M4 sister survival as that of either single mutant (Fig. 3e). By contrast, *ced-9*; *pig-1* double mutants were much more highly penetrant for M4 sister survival than either single mutant (Fig. 3e). These results indicate that *sptf-3* and *ced-9* act in a linear pathway and that *pig-1* and *ced-9*, and hence *pig-1* and the caspase gene *ced-3* (considering that *ced-9* acts by regulating *ced-3*) act in parallel in the regulation of the M4 sister cell death, consistent with previous studies that showed that *pig-1* regulates other apoptotic deaths independently of *ced-3* (refs 4, 5). The *C. elegans* genome encodes three additional caspase genes: *csp-1*, *csp-2* and *csp-3*. We observed that *csp-3*; *csp-1*; *csp-2* triple mutants were not defective in M4 sister cell death (0% of M4 sister survival; *n* = 120). Because *pig-1* mutants are defective in M4 sister cell death, we conclude that *pig-1* promotes apoptosis of the M4 sister independently of *csp-1*, *csp-2* and *csp-3* and hence through a caspase-independent pathway.

We next tested whether *sptf-3* acts through *pig-1* and *egl-1* to promote apoptosis of other cells, namely the AQR sisters, g1A sisters and I2 sisters, all of which survive in *sptf-3* and *pig-1* mutants (Fig. 1f). The wild-type *pig-1* transgene rescued the defect in apoptosis of the AQR sister, g1A sisters and I2 sisters of *pig-1(n4780)* mutants, whereas the *pig-1* transgene containing mutations in the consensus SPTF-3 binding motif (mut.1) did not (Fig. 4a–c). Thus, the consensus SPTF-3 binding motif of the *pig-1* promoter region is necessary for *pig-1* to promote apoptosis of the AQR sister, g1A sisters and I2 sisters, suggesting that SPTF-3 acts through *pig-1* to promote apoptosis of these cells.

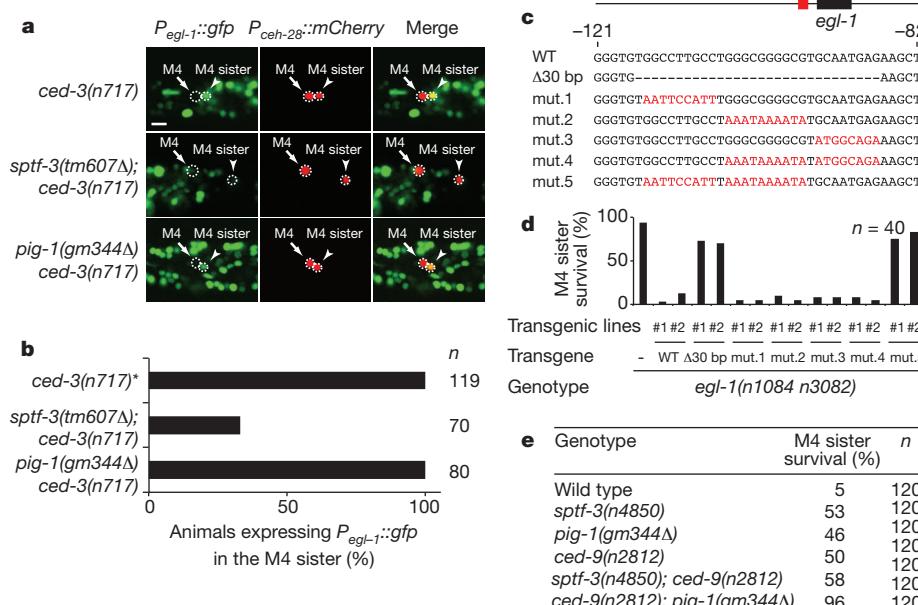


Figure 3 | *sptf-3* directly drives *egl-1* expression in the M4 sister. **a**, *egl-1* expression in the M4 sister cell of animals of the indicated genotypes. **b**, The percentages of animals expressing *P_{egl-1}::gfp* in the M4 sister of the indicated genotypes. *Data from ref. 8. **c**, Representations of the *egl-1* transgenes used in rescue experiments. WT, a wild-type *egl-1* transgene; Δ30 bp lacks 30 base pairs of the SPTF-3-bound region; mut.1 to mut.5, mutated as shown. Black box, *egl-1*-coding region. Red box, SPTF-3-bound region identified by ChIP-seq. Red letters, mutated bases. **d**, The percentages of M4 sister survival in animals of the indicated genotypes. **e**, The percentages of M4 sister survival in animals of the indicated genotypes. All strains were homozygous for *ced-3(n2446)*.

We also tested whether *sptf-3* acts through *egl-1* to promote apoptosis of these cells. Whereas the *egl-1* transgene pTH01 rescued a defect in apoptosis of the AQR sister but not of the g1A sisters or I2 sisters in *egl-1* mutants, the *egl-1* transgene pBC08 (which overlaps only in part with pTH01) rescued the defect in apoptosis of the g1A sisters and I2 sisters but not of the AQR sister of *egl-1* mutants (Supplementary Fig. 9). Both of these *egl-1* transgenes contain the SPTF-3 binding site required for M4 sister cell death. We therefore tested whether the SPTF-3 binding site is required for *egl-1* to promote apoptosis of the AQR sister using pTH01 and of the g1A sisters and I2 sisters using pBC08. The wild-type pTH01 and pBC08 *egl-1* transgenes rescued a defect in apoptosis of the AQR sister, g1A sisters and I2 sisters of *egl-1* mutants (Fig. 4d–f). The mutant pTH01 *egl-1* transgene (mut.5) containing mutations in the SPTF-3 binding site of the *egl-1* promoter region failed to rescue the defect in apoptosis of the AQR sister of *egl-1* mutants, whereas the mutant pBC08 *egl-1* transgene (mut.5) rescued the defect in apoptosis

of the g1A sisters and I2 sisters (Fig. 4d–f). These results indicate that *sptf-3* promotes apoptosis of not only the M4 sister but also of the AQR sister through caspase-dependent and -independent mechanisms through the direct transcriptional activation of *egl-1* and *pig-1*, respectively.

We next determined how *sptf-3* and *pig-1* interact with other genes that specifically regulate M4 sister cell death. We previously reported that the *C. elegans* six family homeodomain protein CEH-34 and the eyes absent protein EYA-1 directly drive the transcription of *egl-1* in the M4 sister to promote M4 sister cell-type-specific apoptosis⁸. *sptf-3(n4850)* synergistically enhanced the M4 sister cell-death defect of *pig-1(gm344Δ)* null mutants, whereas *ceh-34(n4796)* or *eya-1(ok654Δ)* only additively enhanced this defect (Supplementary Table 3). These results indicate that *sptf-3*, *ceh-34* and *eya-1* function in pathways independently of *pig-1*, and that *sptf-3* probably controls *egl-1* at a *cis* regulatory site distinct from that used by the CEH-34-EYA-1 complex in the regulation of M4 sister cell death.

a	Genotype		AQR sister survival (%)	n
Wild type			0	120
<i>pig-1(gm344Δ)</i>			16	120
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #1			0	120
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #2			1	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #1			18	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #2			16	120

b	Genotype		g1A sister survival (%)	n
Wild type			0	240
<i>pig-1(gm344Δ)</i>			16	240
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #1			0	120
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #2			0	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #1			18	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #2			14	120

c	Genotype		I2 sister survival (%)	n
Wild type			0	120
<i>pig-1(gm344Δ)</i>			49	120
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #1			0	120
<i>pig-1(gm344Δ); nEx[pig-1(+)]</i> #2			1	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #1			42	120
<i>pig-1(gm344Δ); nEx[pig-1 mut.1]</i> #2			40	120

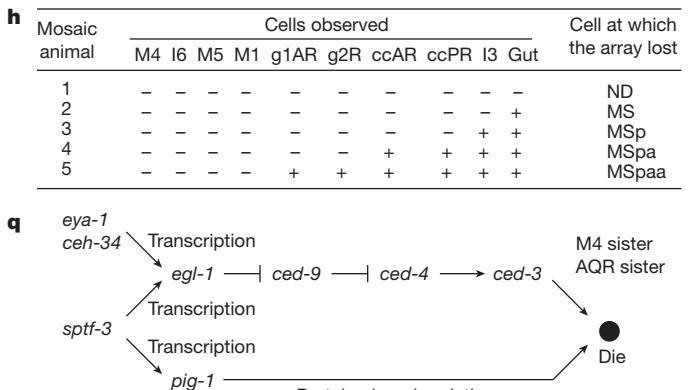
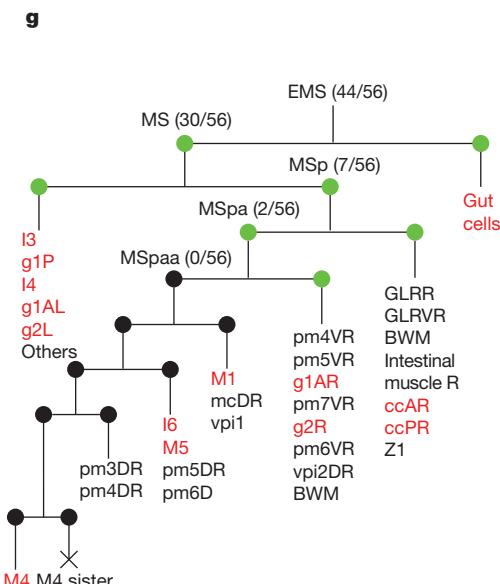
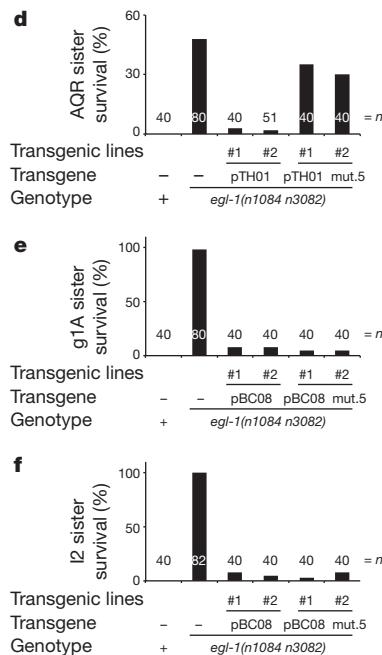


Figure 4 | SPTF-3 functions cell autonomously to promote apoptosis of the M4 sister and the AQR sister. a–f, The percentages of survival of the AQR sister (a, d), g1A sister (b, e) and I2 sister (c, f) in animals of the indicated genotypes. The *pig-1(+)* transgene carries the wild-type *pig-1* genomic locus, and the *pig-1* mut.1 transgene has mutations in the consensus SPTF-3 binding motif in the *pig-1* promoter region, as described in Fig. 2d. The *egl-1* transgene pTH01 contains 6.5 kilobases (kb) of the 5' promoter of *egl-1*, the coding region and 2.2 kb of the 3' of the stop codon of *egl-1*. The *egl-1* transgene pBC08 contains 1.1 kb of the 5' promoter of *egl-1*, the coding region and 5.7 kb of the 3' of the stop codon of *egl-1*. pTH01 mut.5 and pBC08 mut.5 have mutations in the consensus SPTF-3 binding motif of the *egl-1* promoter region, as described

in Fig. 3c. **g,** Diagram of the M4 cell lineage. The cells in red were scored to determine the cell division at which the *sptf-3*-containing extrachromosomal array was lost in a mosaic animal. Shown are numbers of animals carrying the array in the indicated cell of 56 mosaic animals observed. Green circles, blastomeres in which the array was present. Black circles, blastomeres in which the array was lost. **h,** Examples of mosaic animals carrying an extrachromosomal array in the indicated cells. +, extrachromosomal array present. -, extrachromosomal array absent. **i–p,** Nomarski and epifluorescence images of a strain expressing a *GFP::sptf-3* transgene. Scale bar, 10 μm. **q,** A model for the pathways that regulate M4 sister and AQR sister cell-specific programmed cell death.

To identify the cellular site of *sptf-3* action in the regulation of M4 sister cell death, we performed a genetic mosaic analysis (Fig. 4g, h). We generated *sptf-3(n4850)* mutants carrying an extrachromosomal array containing an *sptf-3*-rescuing transgene marked with the cell-autonomous GFP reporters *sur-5:gfp* and *unc-119:gfp* (refs 11 and 12). Extrachromosomal arrays in *C. elegans* are mitotically unstable, resulting in mosaic animals that randomly lose the array in some cell lineages. We observed 56 mosaic animals that were not rescued for the defect in M4 sister cell death but that carried the array with the *sptf-3*-rescuing transgene. Among these 56 animals, we found two animals that retained the array in the blastomere MSpa, the great-great-great-grandmother of the M4 sister (MSpaaaap), but did not find any animals that retained the array in the blastomere MSPaa, the great-great-grandmother of the M4 sister (Fig. 4g, h), indicating that *sptf-3* is required at or later than the stage of the blastomere MSPaa and seems to function cell autonomously to promote the death of the M4 sister.

To determine the temporal and spatial expression pattern of SPTF-3, we generated a *gfp::sptf-3* transgene that encodes functional SPTF-3, as this transgene rescued the M4 sister cell-death defect of *sptf-3* mutants (data not shown). GFP::SPTF-3 was expressed ubiquitously in embryos and early larval animals (Fig. 4i–k and Supplementary Fig. 10); its expression in the M4 neuron and g1A cells decreased during larval development (Fig. 4k, l and Supplementary Fig. 10a, b). GFP::SPTF-3 was strongly expressed in the seam cells and the hyp7 cells (Supplementary Fig. 10c, d), consistent with the observation that overexpression of *sptf-3* induced ectopic expression of *pig-1* in these cells (Fig. 2g, h). GFP::SPTF-3 localized exclusively to nuclei, consistent with the presumed function of SPTF-3 as a transcription factor. We detected GFP::SPTF-3 expression in MSPaa in embryos (Fig. 4i, m) and in the M4 neuron at the first larval stage (Fig. 4k, o), indicating that the SPTF-3 expression pattern is consistent with our conclusion that *sptf-3* functions cell autonomously to promote the death of the M4 sister.

Given the cell-autonomous function of *sptf-3* and *pig-1* (ref. 4), we tested whether expression of *sptf-3* or *pig-1* could induce apoptosis. Ectopic expression of the caspase CED-3 under the control of the *mec-7* promoter caused the PLM neurons to die, whereas similar expression of either SPTF-3 or PIG-1 failed to do so (Supplementary Fig. 11a). Overexpression of SPTF-3 from a multi-copy array under the control of the *sptf-3* promoter did not result in ectopic apoptosis of the seam cells, M4 neuron, AQR neuron, g1A cells and I2 neurons (Supplementary Fig. 11b, c), all of which normally survive. Consistent with these observations, overexpression of SPTF-3 under the control of the *sptf-3* promoter also did not cause ectopic expression of the pro-apoptotic BH3-only gene *egl-1* (Supplementary Fig. 12), embryonic or larval lethality (Supplementary Table 4). These results indicate that expression of neither *sptf-3* nor *pig-1* is sufficient to promote apoptosis and suggest that other factors are probably required for *sptf-3* or *pig-1* to promote apoptosis.

Our findings demonstrate that the apoptosis of the M4 sister cell is specified by at least two parallel pathways that function non-redundantly: (1) the caspase-dependent apoptotic pathway mediated by *egl-1* and activated through direct transcriptional regulation by two different inputs, by *sptf-3* or by *ceh-34* and *eya-1*; and (2) the caspase-independent apoptotic pathway mediated by *pig-1* and activated through direct transcriptional regulation by *sptf-3* (Fig. 4q). In this apoptotic regulatory network, a single transcription factor, SPTF-3, coordinates caspase-dependent and caspase-independent pathways to promote cell-type-specific apoptosis of the M4 sister and the AQR sister.

Our discovery that there is a common regulatory node for caspase-dependent and caspase-independent apoptosis might identify a general mechanism for the regulation of cell-type-specific apoptosis and, if so, could have an important therapeutic impact, as both caspase-dependent and caspase-independent cell-death processes (such as necroptosis) have been implicated in diseases as diverse as glaucoma, heart attacks and neurodegeneration^{13–15}. Our finding that there can be

a common regulatory node for the two pathways reveals a new possible approach to therapeutic intervention.

METHODS SUMMARY

C. elegans strains were cultured at 20 °C as described previously¹⁶. The N2 strain was used as the wild type. The mutations, integrations and extrachromosomal arrays used are described in Methods. *sptf-3(n4850)* and *pig-1(n4780)* were isolated as described⁸. Single nucleotide polymorphisms were used to map *sptf-3(n4850)* and *pig-1(n4780)* to a 136-kb interval (linkage group I: 13,340,839–13,477,415) and a 650-kb interval (linkage group IV: 1–657,968), respectively¹⁷. For ChIP-seq experiments, ChIP was performed as described¹⁸. The enriched DNA and input DNA were used to construct libraries for Illumina/Solexa DNA sequencing using HiSeq 2000. The reads were aligned with the *C. elegans* genome using the software Bowtie and analysed as described in Methods. All SPTF-3-bound regions are described in Supplementary Tables 1 and 2.

Full Methods and any associated references are available in the online version of the paper.

Received 18 September 2012; accepted 30 May 2013.

Published online 14 July 2013.

1. Fuchs, Y. & Steller, H. Programmed cell death in animal development and disease. *Cell* **147**, 742–758 (2011).
2. Yuan, J. & Kroemer, G. Alternative cell death mechanisms in development and beyond. *Genes Dev.* **24**, 2592–2602 (2010).
3. Metzstein, M. M., Stanfield, G. M. & Horvitz, H. R. Genetics of programmed cell death in *C. elegans*: past, present and future. *Trends Genet.* **14**, 410–416 (1998).
4. Cordes, S., Frank, C. A. & Garriga, G. The *C. elegans* MELK ortholog PIG-1 regulates cell size asymmetry and daughter cell fate in asymmetric neuroblast divisions. *Development* **133**, 2747–2756 (2006).
5. Denning, D. P., Hatch, V. & Horvitz, H. R. Programmed elimination of cells by caspase-independent cell extrusion in *C. elegans*. *Nature* **488**, 226–230 (2012).
6. Avery, L. & Horvitz, H. R. A cell that dies during wild-type *C. elegans* development can function as a neuron in a *ced-3* mutant. *Cell* **51**, 1071–1078 (1987).
7. Sulston, J. E., Schierenberg, E., White, J. G. & Thomson, J. N. The embryonic cell lineage of the nematode *Caenorhabditis elegans*. *Dev. Biol.* **100**, 64–119 (1983).
8. Hirose, T., Galvin, B. D. & Horvitz, H. R. Six and Eya promote apoptosis through direct transcriptional activation of the proapoptotic BH3-only gene *egl-1* in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **107**, 15479–15484 (2010).
9. Ulm, E. A., Sleiman, S. F. & Chamberlin, H. M. Developmental functions for the *Caenorhabditis elegans* Sp protein SPTF-3. *Mech. Dev.* **128**, 428–441 (2011).
10. Hengartner, M. O., Ellis, R. E. & Horvitz, H. R. *Caenorhabditis elegans* gene *ced-9* protects cells from programmed cell death. *Nature* **356**, 494–499 (1992).
11. Yochem, J., Gu, T. & Han, M. A new marker for mosaic analysis in *Caenorhabditis elegans* indicates a fusion between *hyp6* and *hyp7*, two major components of the hypodermis. *Genetics* **149**, 1323–1334 (1998).
12. Maduro, M. & Pilgrim, D. Identification and cloning of *unc-119*, a gene expressed in the *Caenorhabditis elegans* nervous system. *Genetics* **141**, 977–988 (1995).
13. Smith, C. C. T. et al. Necrostatin: a potentially novel cardioprotective agent? *Cardiovasc. Drugs Ther.* **21**, 227–233 (2007).
14. Rosenbaum, D. M. et al. Necroptosis, a novel form of caspase-independent cell death, contributes to neuronal damage in a retinal ischemia-reperfusion injury model. *J. Neurosci. Res.* **88**, 1569–1576 (2010).
15. Hyman, B. T. & Yuan, J. Apoptotic and non-apoptotic roles of caspases in neuronal physiology and pathophysiology. *Nature Rev. Neurosci.* **13**, 395–406 (2012).
16. Brenner, S. The genetics of *Caenorhabditis elegans*. *Genetics* **77**, 71–94 (1974).
17. Wicks, S. R., Yeh, R. T., Gish, W. R., Waterston, R. H. & Plasterk, R. H. Rapid gene mapping in *Caenorhabditis elegans* using a high density polymorphism map. *Nature Genet.* **28**, 160–164 (2001).
18. Ercan, S. et al. X chromosome repression by localization of the *C. elegans* dosage compensation machinery to sites of transcription initiation. *Nature Genet.* **39**, 403–408 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Fire, J. Gaudet, C. Barbara and Y. Iino for reporter constructs used to observe cell-type-specific apoptosis; G. Garriga for *pig-1* strains; the *Caenorhabditis* Genetic Center, which is funded by the NIH Office of Research Infrastructure Programs (P40 OD010440) and the National BioResource project for strains; D. Denning, K. Boulias, A. Corrionero and H. Johnsen for comments about the manuscript; and members of the Horvitz laboratory for technical support and discussions. This work was supported by the Howard Hughes Medical Institute. T.H. was supported in part by the Ministry of Education, Science, Technology, Sports and Culture of Japan. H.R.H. is the David H. Koch Professor of Biology at the Massachusetts Institute of Technology and an Investigator of the Howard Hughes Medical Institute.

Author Contributions T.H. and H.R.H. designed the experiments, analysed the data and wrote the manuscript. T.H. performed the experiments.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to H.R.H. (horvitz@mit.edu).

METHODS

C. elegans strains and genetics. *C. elegans* strains were cultured at 20 °C as described previously. The N2 strain was used as the wild type. *sptf-3(n4850)* and *pig-1(n4780)* were isolated as described. Single nucleotide polymorphisms were used to map *sptf-3(n4850)* and *pig-1(n4780)* to a 136-kb interval (linkage group (LG) I: 13,340,839–13,477,415) and a 650-kb interval (LG IV: 1–657,968), respectively. The following mutations, integrations and extrachromosomal arrays were used. LGI: *sptf-3(n4850)*, *tm607A*, *eya-1(ok654A)*, *csp-3(n4872A)*, *nIs177[P_{ceh-28}::gfp, lin-15AB(+)]*, *zdl5s5[P_{mec-4}::gfp, lin-15AB(+)]*. LGII: *csp-1(n4967A)*, *nIs343[P_{egl-1}::gfp, lin-15AB(+)]*. LGIII: *ced-9(n2812)*, *nIs176[P_{ceh-28}::gfp, lin-15AB(+)]*, *nls375[P_{gcy-3}::gfp, lin-15AB(+)]*, *nls390[P_{fp-15}::gfp, lin-15AB(+)]*, *nls579[sptf-3(+), P_{myo-3}::mCherry]*. LGIV: *ced-3(n717, n2446)*, *pig-1(n4780, gm344A)*, *csp-2(n4871A)*, *nIs175[P_{ceh-28}::gfp, lin-15AB(+)]*, *nls578[sptf-3(+), P_{myo-5}::mCherry]*, *wIs78[ajm-1::gfp, scm::gfp, F58E10, unc-119(+)]*. LGV: *ceh-34(n4796)*. LGX: *lin-15(n765)*, *nIs106[P_{lin-11}::gfp, lin-15AB(+)]*, *nIs283[P_{gcy-10}::gfp, lin-15AB(+)]*, *nIs349[P_{ceh-28}::mCherry, lin-15AB(+)]*, *nIs427[P_{phat-1}::gfp, lin-15AB(+)]*, *nIs429[P_{phat-5}::gfp, lin-15AB(+)]*, *nls431[gfp::sptf-3]*, *bcl2s24[P_{tpk-1}::gfp, lin-15AB(+)]*, *nls540[P_{pig-1}::gfp, rol-6(su1006)]*. Unmapped integrations: *nls597[P_{pig-1}::gfp, P_{rab-3}::mCherry, sur-5::mCherry, lin-15AB(+)]*, *nIs601[P_{pig-1} Δ71::gfp, P_{rab-3}::mCherry, sur-5::mCherry, lin-15AB(+)]*, *nIs605[P_{pig-1} mut.1::gfp, P_{rab-3}::mCherry, sur-5::mCherry, lin-15AB(+)]*. Extrachromosomal arrays: *nEx1453* and *nEx1454[sptf-3(+), P_{lin-44}::gfp]*, *nEx1684[sptf-3(+), P_{unc-119}::gfp, sur-5::gfp]*, *nEx1971*, *nEx1972*, *nEx2105*, *nEx2106*, *nEx2109*, *nEx2110*, *nEx2113* and *nEx2114* [*pig-1(+)*, *P_{lin-44}::gfp*], *nEx1974* and *nEx1976* [*pig-1* Δ71 bp, *P_{lin-44}*::gfp], *nEx1978*, *nEx1979*, *nEx2107*, *nEx2108*, *nEx2111*, *nEx2112*, *nEx2115* and *nEx2116* [*pig-1* mut.1, *P_{lin-44}*::gfp], *nEx1998* and *nEx1999* [*pTH01* Δ30 bp, *P_{lin-44}*::gfp], *nEx2000*, *nEx2001*, *nEx2101* and *nEx2102* [*pTH01*, *P_{lin-44}*::gfp], *nEx2002* and *nEx2003* [*pTH01* mut.2, *P_{lin-44}*::gfp], *nEx2004* and *nEx2005* [*pTH01* mut.1, *P_{lin-44}*::gfp], *nEx2006* and *nEx2007* [*pTH01* mut.3, *P_{lin-44}*::gfp], *nEx2008*, *nEx2009*, *nEx2103* and *nEx2104* [*pTH01* mut.5, *P_{lin-44}*::gfp], *nEx2010* and *nEx2011* [*pTH01* mut.4, *P_{lin-44}*::gfp], *nEx2091*, *nEx2092*, *nEx2095*, *nEx2096*, *nEx2099* and *nEx2100* [*pBC08*, *P_{lin-44}*::gfp], *nEx2093*, *nEx2094*, *nEx2098* and *nEx2099* [*pBC08* mut.5, *P_{lin-44}*::gfp], *nEx2117* and *nEx2118* [*p_{ceh-3}::sptf-3(+)*, *P_{mec-5}::mCherry, rol-6(su1006)]*, *nEx2219* and *nEx22120* [*P_{mec-5}::pig-1(+)*, *P_{mec-3}::mCherry, rol-6(su1006)]*, *nEx22121* and *nEx22122* [*P_{mec-5}::ced-3(+)*, *P_{mec-3}::mCherry, rol-6(su1006)]*.

Antibody production. Protein fragments corresponding to amino acids 1–79 and 192–275 of SPTF-3 fused to glutathione S-transferase (GST) were expressed, purified using glutathione Sepharose 4B (Amersham Biosciences) and used to raise SPTF-3 N81 and SPTF-3 M82 antibodies, respectively. Antisera were generated by Pocono Rabbit Farm and Laboratory. Polyclonal antibodies were affinity-purified using identical SPTF-3 fragments fused to maltose-binding protein (MBP) and coupled to Affigel 10 (Bio-Rad).

ChIP-seq analysis and bioinformatic analyses. Chromatin immunoprecipitations were performed as described¹⁸. Wild-type adults were grown on nematode growth medium (NGM) plates, and embryos were obtained by bleaching the gravid adults. Embryos were fixed in 2% formaldehyde for 30 min at room temperature (23 °C), washed once with 100 mM Tris (pH 7.5), twice with M9 buffer and once with FA buffer containing 50 mM HEPES/KOH (pH 7.5), 1 mM EDTA, 1% Triton X-100, 0.1% sodium deoxycholate, 150 mM NaCl with protease inhibitors (Roche), and frozen at -80 °C. In total, 1.7 ml of packed embryos were suspended in 3 ml of FA buffer and homogenized using a dounce homogenizer. The sample was then sonicated using a Branson Sonifier 450 equipped with a microtip in ice water 24 times at the following setting: output control 1.2 and duty cycle constant 30 s ON and 1 min OFF in each cycle. The sample was centrifuged at 13,000g for 15 min at 4 °C. The protein concentration of the embryonic extract was determined using a BCA Protein Assay kit (Thermo Scientific). Sarkosyl was added to the embryonic extract at a final concentration of 1%, and the embryonic extract was centrifuged at 13,000g for 5 min at 4 °C. Embryonic extract containing 1.54 mg of protein was used as the input control. Embryonic extract containing 15.4 mg of protein and 75 µg of an affinity-purified SPTF-3 N81 antibody, embryonic extract containing 15.4 mg of protein and 54 µg of an affinity-purified SPTF-3 M82 antibody or embryonic extract containing 6.4 mg of protein and 54 µg of normal IgG was incubated at 4 °C overnight for immunoprecipitation in 2 ml of FA buffer containing 1% sarkosyl and protease inhibitors. The precipitated immunocomplexes were collected with Dynabeads protein A (Invitrogen) and washed twice with FA buffer for 5 min, once with FA buffer containing 1 M NaCl for 5 min, once with FA buffer containing 500 mM NaCl for 10 min, once with TEL buffer containing 250 mM LiCl, 1% NP-40, 1% sodium deoxycholate, 1 mM EDTA and 10 mM Tris (pH 8.0) for 10 min and twice with TE buffer (pH 8.0) containing 10 mM Tris (pH 8.0) and 1 mM EDTA for 5 min. The immunocomplexes were eluted twice with 300 µl TE (pH 8.0) containing 1% SDS, 250 mM NaCl and TE at 65 °C for 15 min. The eluted and input samples were incubated at 65 °C overnight to reverse the crosslinking and then treated with 67 µg ml⁻¹ of RNase A at room temperature for 1 h followed by 67 µg ml⁻¹ of proteinase K at

55 °C for 2 h. DNA was purified by two phenol-chloroform extractions, followed by precipitation with ethanol and re-suspension in 50 µl of H₂O followed by Illumina/Solexa DNA sequencing using HiSeq 2000.

ChIP-seq data analysis. Images acquired from an Illumina/Solexa sequencer were processed through the bundled Solexa image extraction pipeline, which identified polony positions, performed base-calling and generated QC statistics. ChIP-derived reads were aligned with the *C. elegans* genome, Wormbase WS190, using the software Bowtie. Only sequences that mapped uniquely to the genome with zero or one mismatch were used for further analysis. The enriched sites with P value $< 10^{-9}$ per antibody defined by ChIP-seq were identified as described¹⁹. All SPTF-3-bound regions are described in Supplementary Tables 1 and 2.

SPTF-3 binding motif. The sequences for all regions precipitated with an SPTF-3 M82 antibody were analysed using a motif-based sequence analysis tool for large DNA data sets, Multiple Expectation Maximization for Motif Elicitation (MEME)-ChIP²⁰ (<http://meme.sdsc.edu/meme/cgi-bin/meme-chip.cgi>) to identify SPTF-3 binding-site motifs. Motifs between six and 30 nucleotides were considered, with a maximum of six motifs for a data set input. The motifs discovered by MEME-ChIP were input to MAST²¹ to determine consensus motifs from the MEME-ChIP output and to search a sequence database for sequences that match the motifs.

Gene Ontology analysis. A total of 2,459 SPTF-3-enriched binding regions were identified in ChIP experiments using both the anti-SPTF-3 N81 antibody and anti-SPTF-3 M82 antibodies. Putative target genes with overlapping SPTF-3 binding sites located upstream (≤ 10 kb from the transcription start site) or within the genes were identified, and these genes were subjected to a Gene Ontology enrichment analysis using GStat²² (<http://gostat.wehi.edu.au/>). The complete set of annotated genes was used as the background set of genes.

Plasmid construction. The $P_{ceh-28::gfp}$, $P_{ceh-28::mCherry}$, $P_{gcy-10::gfp}$, $P_{egl-1::gfp}$ and wild-type $egl-1$ (pTH01 or pBC08) transgenes were described previously^{8,23}. The $sptf-3$ transgene contained 2.0 kb of 5' promoter, the coding region and 0.5 kb 3' of the stop codon in the pGEM-T Easy vector. The $gfp::sptf-3$ transgene contained 2.0 kb of the 5' promoter of $sptf-3$, the gfp gene with synthetic introns, the coding region and 0.5 kb 3' of the stop codon of $sptf-3$ in pRS426. The $pig-1$ genomic fragment containing 0.8 kb of 5' promoter, the coding region and 0.6 kb 3' of the $pig-1$ stop codon was cloned into pRS426, and the last intron was removed using two NruI restriction enzyme sites. The QuickChange II XL Site-Directed Mutagenesis Kit (Stratagene) was used to generate the $pig-1(\Delta 71\text{ bp})$ construct lacking the SPTF-3-bound region, the $pig-1\text{ mut.1}$ construct mutated in the consensus SPTF-3 binding motif, the $egl-1(\Delta 30\text{ bp})$ construct and the $egl-1\text{ mut.1-mut.5}$ constructs. $sptf-3$ complementary DNA was isolated by RT-PCR. The $sptf-3$ cDNA fragment corresponding to amino acids 1–79 or 192–275 was cloned in pGEX-4T-3 (GE Healthcare Life Sciences) and pMAL-c2 (New England BioLabs) to express SPTF-3 protein fragments fused with GST or MBP, respectively. The $phat-5$ promoter sequence in pGD48, provided from J. Gaudet, was cloned in pPD122.56 to generate the $P_{phat-5::gfp}$ transgene. The $P_{flp-15::gfp}$ transgene contained 2.4 kb of the 5' promoter of $flp-15$ in pPD122.56. The $P_{gcy-37::gfp}$ transgene contained 1.1 kb of the 5' promoter of $gcy-37$ in pPD122.56. For the $P_{mec-z::sptf-3}$, $P_{mec-z::pig-1}$ and $P_{mec-z::ced-3}$ transgenes, cDNAs containing the entire coding region of each gene were cloned in pPD96.41. The $P_{pig-1::gfp}$, $P_{pig-1(\Delta 71\text{ bp})::gfp}$ and $P_{pig-1\text{ mut.1}::gfp}$ transgenes contained 0.9 kb of the 5' promoter of $pig-1$ in pPD122.56. The specific primer sequences are available on request from the authors.

Germline transformation. Germline transformation experiments were performed as described²⁴. The *gfp* or *mCherry* transgenes were injected at 50 or 100 $\mu\text{g ml}^{-1}$ into *lin-15(n765ts)* or *ced-3(n717)lin-15(n765ts)* animals with 50 $\mu\text{g ml}^{-1}$ of pL15EK as a co-injection marker²⁵. The *sptf-3* transgene was injected at 10 $\mu\text{g ml}^{-1}$ into *sptf-3(n4850)* animals with 50 $\mu\text{g ml}^{-1}$ of *P_{lin-44::gfp}* as a co-injection marker²⁶ to rescue the defect in M4 sister cell death. For expression of *sptf-3*, the *sptf-3* transgene was injected at 50 $\mu\text{g ml}^{-1}$ into *nIs540* animals with 5 $\mu\text{g ml}^{-1}$ of *P_{myo-3::mCherry}* as a co-injection marker²⁷.

Quantitative RT-PCR. Total RNA from wild-type *sptf-3(n4850); nls349* embryos was prepared using an RNeasy Mini kit (Qiagen). Reverse transcription and quantitative PCR were performed as described²⁸. The data presented were generated from three PCR reactions, and *rpl-26* mRNA levels were used for normalization. The specific primer sequences are available on request from the authors.

RNAi analysis. Nucleotides 629–1,238 of the *sptf-3* cDNA and nucleotides 694–2,112 of the *pig-1* cDNA were cloned into the pBluescript II vector, respectively. Sense and antisense RNA molecules were synthesized using T3 and T7 RNA polymerase, respectively, and then annealed to generate double-stranded RNA. The double-stranded RNA was injected into the gonads of *nls175* adult hermaphrodites, and their progeny were scored at the first larval stage for a defect in M4 sister cell death.

Mosaic analysis. Transgenic animals of genotype *sptf-3(n4850); nIs349; nEx 1684[sptf-3(+), P_{unc-119::gfp}, sur-5::gfp]* were generated for the mosaic analysis

experiments. Fifty-six mosaic animals that carried the extrachromosomal array but that were not rescued for the defect in M4 sister cell death were picked and observed using Nomarski optics and epifluorescence to determine the presence or absence of the extrachromosomal array as judged by GFP fluorescence in the following cells: M4, I5, M5, M1, g1AR, g2R, ccAR, ccPR, I3, g1P, I4, g1AL, g2L and intestinal cells. The cell division at which each extrachromosomal array was lost was determined based on the cells that retained the array⁷.

Analyses of defects in programmed cell deaths of specific cells. The programmed cell deaths of specific cells were scored at the indicated stages of larval development using the following transgenes, which express GFP in specific cells: M4 sister cell death, *nls175*, *nls176* or *nls177* at the L1 stage; g1A sister cell death, *nls429* at the L1 stage; I2 sister cell death, *nls390* at the L4 stage; AQR sister cell death, *nls375* at the L2 stage; NSM sister cell death, *bcls24* at the L1 stage; II sister cell death, *nls283* at the L4 stage; VC homologue cell deaths, *nls106* at the L4 stage. The number of the V-lineage-derived seam cells of young adult animals was scored using *wIs78*, which expresses GFP in the seam cells. A fluorescence-equipped compound microscope was used to score these programmed cell deaths.

Identification of the M4 neuron and the MSpaa cell. The M4 neuron was identified on the basis of its distinctive position in the anterior pharynx. To identify the MSpaa cell, we traced its cell lineage starting from the four-cell-stage embryo, at which point it is possible to easily distinguish the four blastomeres ABa, ABp, P2 and EMS (the progenitor of MSpaa).

19. Marson, A. et al. Connecting microRNA genes to the core transcriptional regulatory circuitry of embryonic stem cells. *Cell* **134**, 521–533 (2008).
20. Bailey, T. L. MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.* **37**, W202–W208 (2009).
21. Bailey, T. L. & Gribskov, M. Combining evidence using p-values: application to sequence homology searches. *Bioinformatics* **14**, 48–54 (1998).
22. Beissbarth, T. & Speed, T. P. GOstat: find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20**, 1464–1465 (2004).
23. Conradt, B. & Horvitz, H. R. The *C. elegans* protein EGL-1 is required for programmed cell death and interacts with the Bcl-2-like protein CED-9. *Cell* **93**, 519–529 (1998).
24. Mello, C. C., Kramer, J. M., Stinchcomb, D. & Ambros, V. Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. *EMBO J.* **10**, 3959–3970 (1991).
25. Clark, S. G., Lu, X. & Horvitz, H. R. The *Caenorhabditis elegans* locus *lin-15*, a negative regulator of a tyrosine kinase signalling pathway, encodes two different proteins. *Genetics* **137**, 987–997 (1994).
26. Herman, M. A., Vassilieva, L. L., Horvitz, H. R., Shaw, J. E. & Herman, R. K. The *C. elegans* gene *lin-44*, which controls the polarity of certain asymmetric cell divisions, encodes a Wnt protein and acts cell nonautonomously. *Cell* **83**, 101–110 (1995).
27. Frøkjær-Jensen, C. et al. Single-copy insertion of transgenes in *Caenorhabditis elegans*. *Nature Genet.* **40**, 1375–1383 (2008).
28. Andersen, E. C., Saffer, A. M. & Horvitz, H. R. Multiple levels of redundant processes inhibit *Caenorhabditis elegans* vulval cell fates. *Genetics* **179**, 2001–2012 (2008).

Epithelial junctions maintain tissue architecture by directing planar spindle orientation

Yu-ichiro Nakajima¹, Emily J. Meyer¹, Amanda Kroesen¹, Sean A. McKinney¹ & Matthew C. Gibson^{1,2}

During epithelial cell proliferation, planar alignment of the mitotic spindle coordinates the local process of symmetric cell cleavage with the global maintenance of polarized tissue architecture^{1,2}. Although the disruption of planar spindle alignment is proposed to cause epithelial to mesenchymal transition and cancer^{3–6}, the *in vivo* mechanisms regulating mitotic spindle orientation remain elusive. Here we demonstrate that the actomyosin cortex and the junction-localized neoplastic tumour suppressors Scribbled and Discs large 1 have essential roles in planar spindle alignment and thus the control of epithelial integrity in the *Drosophila* imaginal disc. We show that defective alignment of the mitotic spindle correlates with cell delamination and apoptotic death, and that blocking the death of misaligned cells is sufficient to drive the formation of basally localized tumour-like masses. These findings indicate a key role for junction-mediated spindle alignment in the maintenance of epithelial integrity, and also reveal a previously unknown cell-death-mediated tumour-suppressor function inherent in the polarized architecture of epithelia.

Throughout the animal kingdom, mitotic division typically initiates with cell rounding driven by Rho-kinase (ROK)-dependent actomyosin contractility at the cell cortex^{7,8}. In polarized epithelia, this process is generally followed by planar alignment of the mitotic spindle and cell cleavage orthogonal to the plane of the epithelium^{1,7,9–14}. These events effectively coordinate the geometry of chromosome segregation with the architecture of polarized cell–cell junctions, thereby ensuring continuous epithelial integrity. Nevertheless, despite the critical nature of

this function, the mechanisms that restrict spindle orientation to the plane of the epithelium remain poorly understood, as do the precise consequences of spindle misalignment for epithelial organization.

In order to uncover the mechanism guiding planar orientation of the mitotic spindle in the *Drosophila* wing disc epithelium, we first analysed spindle pole movements with a fluorescent centriolar marker, Asterless-yellow fluorescent protein (ASL-YFP), or a Centrosomin-green fluorescent protein (CNN-GFP) fusion protein (Fig. 1a and Supplementary Fig. 1b, c). In interphase cells, centrioles marked with ASL-YFP consistently localized in a super-apical position, well above the plane of the adherens junctions or basolateral septate junctions (Supplementary Fig. 1a, b). At prophase entry, centrosomes moved basally (Fig. 1b and Supplementary Fig. 1c), and the mitotic spindle aligned with the plane of the epithelium during metaphase (averaging $6.5 \pm 4.9^\circ$ from planar) and anaphase ($4.6 \pm 3.7^\circ$ from planar; Fig. 1a, b and Supplementary Fig. 1d, e). Indeed, mitotic centrosomes were almost always constrained to the region of the cell cortex defined by NRG-GFP accumulation at the septate junctions ($n = 97$ out of 100 metaphase figures; Supplementary Fig. 1d). Accordingly, metaphase and anaphase spindle orientations were consistently within the maximal limits set by the septate junction-delimited mitotic zone (estimated at $25.6 \pm 1.3^\circ$, $n = 91$).

The mitotic spindle could achieve planar orientation through either an active mechanism involving spatial cues or, more simply, alignment with the cellular long axis (for example, Hertwig's rule). Inconsistent with a long-axis orientation mechanism, however, both interphase and

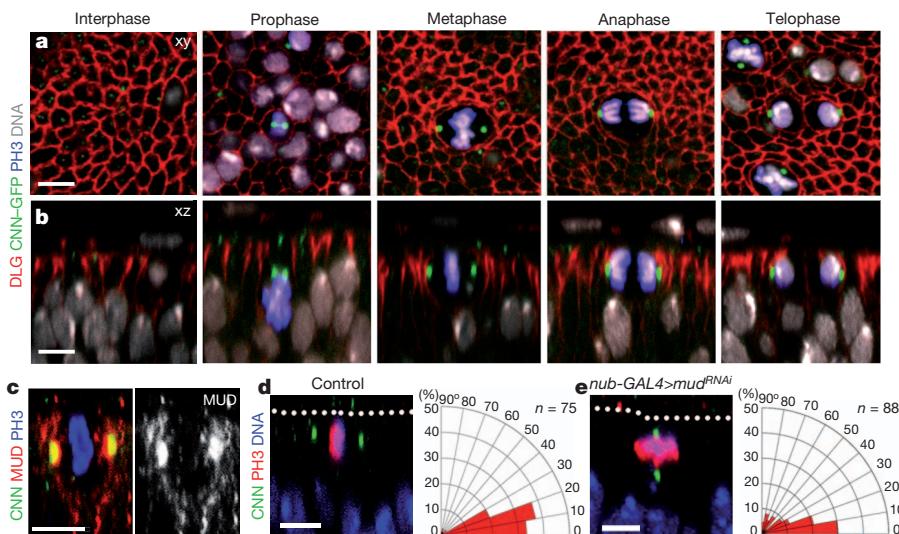


Figure 1 | Planar orientation of the mitotic spindle during wing disc development. **a, b**, Images of mitotic spindles in xy (a) and xz (b) images. **c**, In metaphase, MUD localized to the cell cortex and was enriched on the spindle poles. Left panel is a merged image of CNN, PH3 to label mitotic chromosomes,

and MUD. Right panel shows MUD alone. **d, e**, Quantification of mitotic spindle alignment in control (d) and *mud* RNAi (e) wing discs. n indicates the number of spindles observed. Scale bars, 5 μ m.

¹Stowers Institute for Medical Research, 1000 East 50th Street, Kansas City, Missouri 64110, USA. ²Department of Anatomy and Cell Biology, University of Kansas Medical Center, 3901 Rainbow Boulevard, Kansas City, Kansas 66160, USA.

mitotic cells were more extended along their apicobasal axes (data not shown; Supplementary Fig. 1f), suggesting that molecular cues guide planar spindle alignment. MUD, a *Drosophila* orthologue of the vertebrate spindle regulator nuclear mitotic apparatus protein (NuMA), controls spindle orientation in neuroblasts and embryonic epithelia^{15–17}. In wing discs, MUD was strongly enriched on the spindle poles during mitosis (Fig. 1c and Supplementary Fig. 2a). To determine the role of MUD in planar spindle orientation, we expressed an RNA interference (RNAi) knockdown construct under the control of the wing-specific driver *nubbin-GAL4* (*nub-GAL4>UAS-mud^{RNAi}*). In *mud* RNAi discs, 36.4% of the spindles exhibited aberrant planar orientation ($n = 88$; Fig. 1d, e and Supplementary Fig. 2b, c), indicating that MUD functions to position the spindle in the plane of the epithelium.

During asymmetric cell divisions in *Drosophila* and mammalian tissue culture, cues embedded in the actomyosin cortex direct spindle orientation through interaction with astral microtubules^{18,19}. In wing discs, we observed a pronounced accumulation of F-actin, phospho-myosin and the actin-binding protein moesin (MOE-GFP) at the cortex of mitotic cells (Fig. 2a, b)⁷. MOE is the sole *Drosophila* ezrin/radixin/moesin (ERM) protein, and upon phospho-activation it cross-links the plasma membrane to the actomyosin cortex to modulate cell shape²⁰. Consistent with a potential role for MOE in both mitotic rounding and planar spindle orientation in the wing disc, phospho-MOE was enriched in mitotic cells, where it localized in a polarized manner to the basal cortex (Fig. 2c, d and Supplementary Fig. 3a).

To test the role of the cell cortex in planar spindle alignment, we used cytochalasin D (CytoD) to disrupt actin polymerization and Y-27632 to disrupt ROK, which phospho-activates both myosin and MOE^{20,21}. Notably, following a 30-min incubation of live wing discs with CytoD, mitotic spindles rotated orthogonal to the plane of the epithelium (78.4% strongly misoriented, $n = 111$; Fig. 2e, f and Supplementary Fig. 3b). Inhibition of ROK with Y-27632 caused similar defects, with 80.8% of spindles becoming severely misoriented ($n = 26$; Fig. 2e, g and Supplementary Fig. 3b). To confirm these results genetically, we knocked down *rok* and *Moe* by RNAi. In discs expressing *rok* RNAi (*nub-GAL4>UAS-rok^{RNAi}*), 58.6% of spindles were misoriented orthogonal to the plane of the epithelium, and another 19.5% exhibited milder defects ($n = 133$; Fig. 2h, i and Supplementary Fig. 3c). We also expressed *rok* RNAi in cell clones induced by the flp-out/GAL4

technique and observed similar abnormalities (Supplementary Fig. 3e). Consistent with these effects, 28.6% of spindles from *nub-GAL4>Moe^{RNAi}* were strongly misoriented and 17.9% had milder abnormalities ($n = 112$; Fig. 2j and Supplementary Fig. 3c). These phenotypes were confirmed with independent constructs (Supplementary Fig. 3d), indicating that ROK and MOE function to control planar spindle orientation during mitosis.

We next investigated how the plane of the epithelium is interpreted by mitotic cells. Septate junctions, which define the mitotic zone and correlate with the positioning of the spindle poles (Fig. 1), correspond to vertebrate tight junctions, and are composed of core components including Coracle, Neurexin IV and Neurogian, as well as the associated neoplastic tumour suppressors Scribbled (SCRIB) and Discs large 1 (DLG, also known as DLG1) (Fig. 3a and Supplementary Fig. 4a–d)²². Intriguingly, mitotic spindle poles localized within 100 nm of septate junctions in intact wing discs (Fig. 3b, c and Supplementary Fig. 4e). Further, extensive epithelial disorganization is observed in *scrib* and *dlg* mutant imaginal discs²³, raising the possibility that their neoplastic phenotypes partly reflect defects in the control of planar spindle alignment.

To determine the requirements for SCRIB and DLG in mitotic spindle orientation, we first used RNAi to knock down *scrib* in developing wing discs (*nub-GAL4>UAS-scrib^{RNAi}*; Supplementary Fig. 4c, d). Unexpectedly, given its role in the establishment of epithelial polarity²⁴, cells expressing *scrib* RNAi exhibited no obvious apicobasal polarity defects in the late third instar (Fig. 3d and Supplementary Fig. 5). At the same experimental time point, however, more than 40% of spindles showed abnormal planar orientation (Fig. 3d–f and Supplementary Fig. 6a). The frequency of spindle misorientation progressively increased during an abnormal developmental delay (Supplementary Fig. 6b, c). Consistent with cooperative activity of SCRIB and DLG, DLG localization became diffuse after *scrib* knockdown (Supplementary Fig. 7a), and we observed spindle orientation defects in discs expressing *dlg* RNAi (Fig. 3g and Supplementary Fig. 7b, c). To confirm these results, we generated mutant clones using the mosaic analysis with a repressible cell marker (MARCM) technique. Spindle orientation was nearly randomized in *scrib¹* mutant cells undergoing mitosis at the apical epithelial surface (Fig. 3h, i and Supplementary Fig. 8a–c). This contrasts with the tendency for the spindle to reorient perpendicular to the plane of the epithelium following disruption of actin or ROK (Fig. 2e–j),

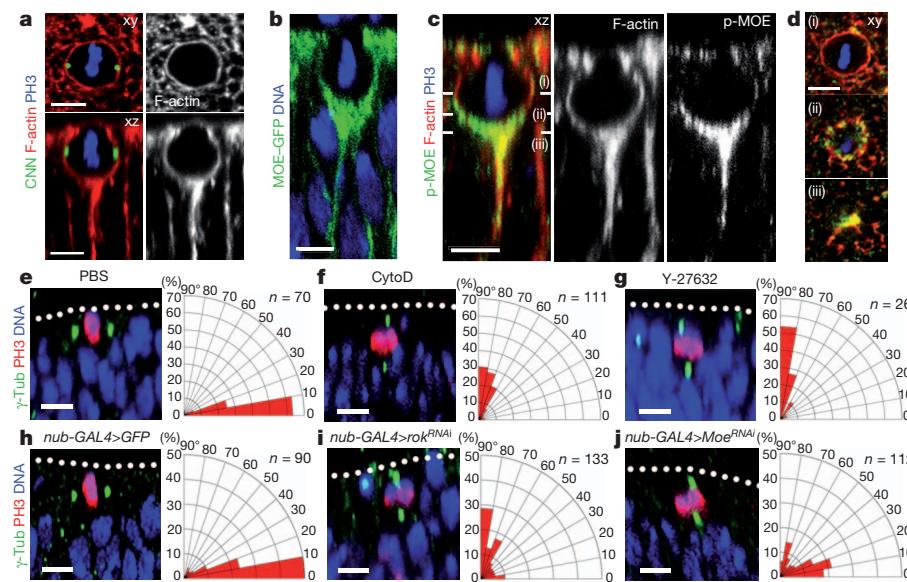


Figure 2 | The actomyosin cortex is required for planar spindle orientation. **a, b**, F-actin accumulated at the cortex of mitotic cells. **b**, MOE-GFP localized to the cortex of mitotic cells in the presumptive wing pouch. **c, d**, Phospho-MOE (p-MOE) was specifically enriched in the basal cortex of mitotic cells. Panels in **d** show different planes according to the z-positions indicated in **c**. **e–g**, Quantification of mitotic spindle alignments in discs treated with PBS (control) (**e**), CytoD (**f**) and Y-27632 (**g**) for 30 min. **h–j**, Quantification of mitotic spindle alignment in wing discs expressing GFP (control) (**h**), *rok^{RNAi}* (**i**) and *Moe^{RNAi}* (**j**). n indicates the number of spindles observed. γ -Tub, γ -tubulin. Scale bars, 5 μ m.

technique and observed similar abnormalities (Supplementary Fig. 3e). Consistent with these effects, 28.6% of spindles from *nub-GAL4>Moe^{RNAi}* were strongly misoriented and 17.9% had milder abnormalities ($n = 112$; Fig. 2j and Supplementary Fig. 3c). These phenotypes were confirmed with independent constructs (Supplementary Fig. 3d), indicating that ROK and MOE function to control planar spindle orientation during mitosis.

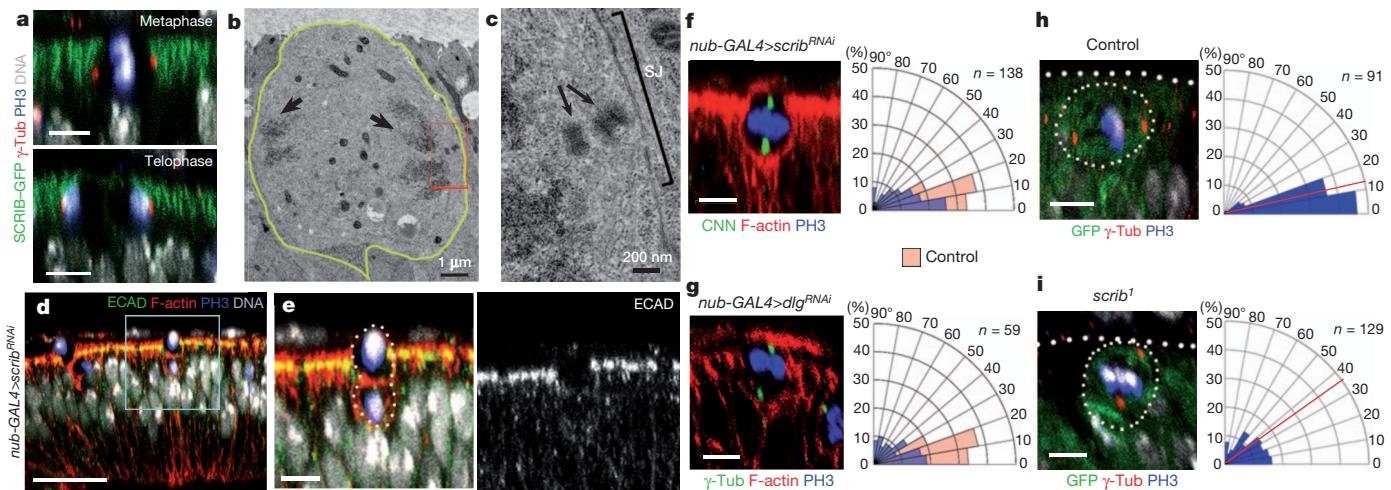


Figure 3 | SCIRB and DLG determine planar orientation of the mitotic spindle. **a**, SCIRB-GFP localized to septate junctions, correlating with spindle position. **b**, A rounded anaphase cell (yellow line) exhibits chromosome segregation (arrows). **c**, Centrioles (arrows) localized in close proximity to septate junctions (SJ, bracket) in a magnification of the red box from **b**. **d**, **e**, *nub-GAL4>UAS-scrib^{RNAi}* wing discs exhibit misoriented spindles (dotted line in **e**) despite normal epithelial architecture and normal apical

localization of ECAD. **e**, Magnification of the box in **d**. **f**, **g**, Quantification of mitotic spindle alignments in *scrib* RNAi (**f**) and *dlg* RNAi (**g**) wing discs compared with control data from Fig. 1d. **h**, **i**, Quantification of metaphase spindle alignments in control and *scrib¹* MARCM clones. The red line shows the median angular deviation. *n* indicates the number of spindles observed. Scale bars, 5 μm (**a**, **e–i**) and 20 μm (**d**).

perhaps owing to defects in mitotic cell rounding in those manipulations (Supplementary Fig. 9). Taken together, these results show that SCIRB and DLG function as determinants of planar spindle orientation and raise the possibility that the neoplastic phenotypes associated with lesions in *scrib* or *dlg* partly result from spindle orientation defects (Fig. 3d, e and Supplementary Fig. 6). Hypothetically, a primary defect in spindle orientation could cause a secondary breakdown of junctional architecture, leading to what is effectively an epithelial-to-mesenchymal transition (EMT).

Loss of tissue architecture is a hallmark of epithelial cancer²⁵, and our results indicate that defects in planar spindle orientation could be a critical initiating event in neoplastic overgrowth of *scrib* and *dlg* mutant discs (Supplementary Fig. 6c). However, misorientation of the spindle may not be sufficient to initiate tumorigenesis, as neoplastic phenotypes were not observed after *mud* or *rok* knockdown. Instead, in *nub-GAL4>rok^{RNAi}* discs and in clones expressing *rok* RNAi, basally localized cells lost polarity markers, underwent apoptotic cell death, and were removed from the epithelium, as previously observed in *Moe* mutants (Fig. 4a, b and Supplementary Fig. 10a)²⁶. To determine the consequence of spindle misorientation, we established a custom selective plane illumination microscopy (SPIM) system. In *nub-GAL4>rok^{RNAi}* wing discs, the basal daughters of severely misoriented divisions exhibited rapid basal movements, consistent with their eventual delamination from the epithelium (Supplementary Fig. 10d–f and Supplementary Videos 1, 2). Further, basally delaminated cells frequently exhibited centrosomal accumulation of CNN-GFP in the absence of mitotic markers (Supplementary Fig. 10g–i), suggesting that these cells may have undergone abnormal divisions and delaminated from the epithelium before their programmed death.

To block apoptosis of misaligned daughter cells in *nub-GAL4>rok^{RNAi}* discs, we co-expressed the caspase inhibitor *p35*. This resulted in a highly disorganized epithelial structure when compared to *nub-GAL4>rok^{RNAi}* alone (Fig. 4c, d). Intriguingly, increased numbers of apoptotic cells were also observed following *scrib* knockdown (Supplementary Fig. 10b), and the neoplastic phenotype was again strongly enhanced by caspase inhibition (Supplementary Fig. 11a). Combined, these results suggest that abnormal epithelial cells arising from spindle misalignment are normally cleared by apoptotic cell death and thereby prevented from disrupting epithelial organization.

To extend these findings as a model for EMT by aberrant spindle orientation, we knocked down *rok* in cell clones co-expressing *UAS-p35*.

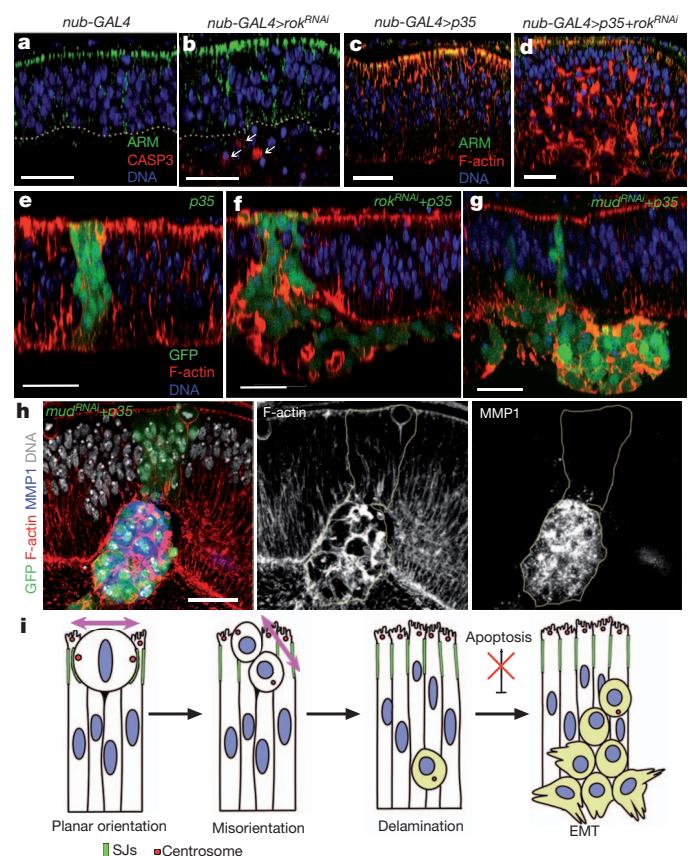


Figure 4 | In the absence of apoptosis, spindle misorientation induces EMT-like effects. **a**, **b**, *rok* knockdown induced basal cell delamination, loss of the Armadillo polarity marker (ARM), and frequent apoptosis. **c**, **d**, Suppression of cell death with *p35*-enhanced tissue disorganization in *rok* RNAi wing discs. **e**, Control clones expressing *p35* were normally integrated in the epithelial layer. By contrast, GFP⁺ cell clones expressing *p35* with *rok* RNAi (**f**) or *mud* RNAi (**g**, **h**) exhibited morphological and molecular characteristics of EMT. **h**, Strong MMP1 expression and actin remodelling observed in a basal tumour-like mass. **i**, Conceptual model for the contribution of spindle misorientation to aberrant EMT. Scale bars, 20 μm.

These clones formed large disorganized masses of aberrant cells on the basal side of the disc ($n = 55$ out of 55 discs, Fig. 4e, f). Intriguingly, the abnormal cells exhibited molecular characteristics of EMT²⁷, including loss of E-cadherin, induction of Matrix metalloproteinase 1 (MMP1) and actin remodelling (Supplementary Fig. 12a, c, e). A similar EMT-like effect was observed in *scrib* RNAi clones co-expressing *p35* ($n = 40$ out of 61 discs; Supplementary Fig. 11b), as well as in *scrib*^l mutant cells expressing *UAS-p35* (Supplementary Fig. 11c). Finally, to more directly implicate spindle misorientation in this process, we used *UAS-p35* to suppress the cell death normally observed in cells expressing *mud* RNAi (Supplementary Fig. 10c). This manipulation phenocopied the effects of disrupting *rok* or *scrib* ($n = 33$ out of 38 discs; Fig. 4g, h and Supplementary Fig. 12b, d, f), demonstrating that in the absence of corrective cell death, spindle misorientation is alone sufficient to induce many key features of EMT. Crucially, the inhibition of apoptosis in dying cells did not lead to the formation of basal tumour-like masses, indicating that the EMT-like effect cannot simply be attributed to forced cell survival ($n = 1$ out of 54 discs; Supplementary Fig. 11d). Combined, these results show that misorientation of the mitotic spindle is typically corrected by cell death, but can result in EMT under conditions where normal apoptotic mechanisms are compromised (Fig. 4i).

Here we demonstrate that planar alignment of the mitotic spindle requires interactions between the mitotic apparatus, the actomyosin cortex and the septate-junction-localized scaffolding proteins SCRIB and DLG. Although SCRIB and DLG have long been implicated in the suppression of neoplastic tumours in *Drosophila* and humans^{23,28,29}, our results uncover a novel requirement for both proteins during alignment of the mitotic spindle to the plane of the epithelium (Fig. 3). By contrast, other potential tumour suppressors linked to cell polarity did not affect planar spindle orientation in the wing disc (Supplementary Table 1). On the basis of these results, we propose that aberrant cleavage plane orientation could be a significant contributor to the loss of epithelial integrity in *scrib* or *dlg* mutant cells (Fig. 4i).

In addition to the above findings, our results demonstrate a critical function for planar orientation of the mitotic spindle in the maintenance of epithelial architecture. Consistent with this view, disruption of spindle orientation with *rok* or *mud* RNAi was sufficient to trigger delamination of basal daughter cells (Supplementary Fig. 10 and Supplementary Videos 1, 2). When combined with the suppression of cell death, these manipulations produced delaminated mesenchyme-like masses exhibiting some molecular characteristics of EMT (Fig. 4h and Supplementary Fig. 12). During the early steps of tumour progression in humans, sporadic mutant cells are thought to acquire a potential to outgrow normal tissue by escaping from the epithelial sheet through EMT or EMT-like events^{5,30}. On the basis of our findings, we propose that the deleterious effects of aberrant spindle alignment are typically corrected by apoptosis, and that suppression of this corrective mechanism could be a common initial driver of epithelial dysplasia and tumorigenesis *in vivo*.

METHODS SUMMARY

Detailed information including the fly stocks, immunofluorescence, drug treatment, imaging and image analysis, and associated references, are included in Methods.

Full Methods and any associated references are available in the online version of the paper.

Received 24 October 2012; accepted 24 May 2013.

Published online 21 July 2013.

1. Morin, X. & Bellaïche, Y. Mitotic spindle orientation in asymmetric and symmetric cell divisions during animal development. *Dev. Cell* **21**, 102–119 (2011).
2. Gillies, T. E. & Cabernard, C. Cell division orientation in animals. *Curr. Biol.* **21**, R599–R609 (2011).
3. Noatynska, A., Gotta, M. & Meraldi, P. Mitotic spindle (DIS)orientation and DISease: cause or consequence? *J. Cell Biol.* **199**, 1025–1035 (2012).

4. Pease, J. C. & Tirnauer, J. S. Mitotic spindle misorientation in cancer—out of alignment and into the fire. *J. Cell Sci.* **124**, 1007–1016 (2011).
5. Vasiliev, J. M., Omelchenko, T., Gelfand, I. M., Feder, H. H. & Bonder, E. M. Rho overexpression leads to mitosis-associated detachment of cells from epithelial sheets: a link to the mechanism of cancer dissemination. *Proc. Natl. Acad. Sci. USA* **101**, 12526–12530 (2004).
6. Fleming, E. S., Temchin, M., Wu, Q., Maggio-Price, L. & Tirnauer, J. S. Spindle misorientation in tumors from *APC*_{min/+} mice. *Mol. Carcinog.* **48**, 592–598 (2009).
7. Meyer, E. J., Ikmi, A. & Gibson, M. C. Interkinetic nuclear migration is a broadly conserved feature of cell division in pseudostratified epithelia. *Curr. Biol.* **21**, 485–491 (2011).
8. Spear, P. C. & Erickson, C. A. Apical movement during interkinetic nuclear migration is a two-step process. *Dev. Biol.* **370**, 33–41 (2012).
9. Reinsch, S. & Karsenti, E. Orientation of spindle axis and distribution of plasma membrane proteins during cell division in polarized MDCKII cells. *J. Cell Biol.* **126**, 1509–1526 (1994).
10. Gibson, M. C., Patel, A. B., Nagpal, R. & Perrimon, N. The emergence of geometric order in proliferating metazoan epithelia. *Nature* **442**, 1038–1041 (2006).
11. Fleming, E. S. et al. Planar spindle orientation and asymmetric cytokinesis in the mouse small intestine. *J. Histochim. Cytochem.* **55**, 1173–1180 (2007).
12. Lu, B., Roegiers, F., Jan, L. Y. & Jan, Y. N. Adherens junctions inhibit asymmetric division in the *Drosophila* epithelium. *Nature* **409**, 522–525 (2001).
13. Ségalen, M. et al. The Fz-Dsh planar cell polarity pathway induces oriented cell division via Mud/NuMA in *Drosophila* and zebrafish. *Dev. Cell* **19**, 740–752 (2010).
14. Guilgur, L. G., Prudêncio, P., Ferreira, T., Pimenta-Marques, A. R. & Martinho, R. G. *Drosophila* aPKC is required for mitotic spindle orientation during symmetric division of epithelial cells. *Development* **139**, 503–513 (2012).
15. Siller, K. H., Cabernard, C. & Doe, C. Q. The NuMA-related Mud protein binds Pins and regulates spindle orientation in *Drosophila* neuroblasts. *Nature Cell Biol.* **8**, 594–600 (2006).
16. Izumi, Y., Ohta, N., Hisata, K., Raabe, T. & Matsuzaki, F. *Drosophila* Pins-binding protein Mud regulates spindle-polarity coupling and centrosome organization. *Nature Cell Biol.* **8**, 586–593 (2006).
17. Bowman, S. K., Neumüller, R. A., Novatchkova, M., Du, Q. & Knoblich, J. A. The *Drosophila* NuMA homolog Mud regulates spindle orientation in asymmetric cell division. *Dev. Cell* **10**, 731–742 (2006).
18. Siller, K. H. & Doe, C. Q. Spindle orientation during asymmetric cell division. *Nature Cell Biol.* **11**, 365–374 (2009).
19. Sandquist, J. C., Kita, A. M. & Bement, W. M. And the dead shall rise: actin and myosin return to the spindle. *Dev. Cell* **21**, 410–419 (2011).
20. Fehon, R. G., McClatchey, A. I. & Bretscher, A. Organizing the cell cortex: the role of ERM proteins. *Nature Rev. Mol. Cell Biol.* **11**, 276–287 (2010).
21. Quintin, S., Gally, C. & Labouesse, M. Epithelial morphogenesis in embryos: asymmetries, motors and brakes. *Trends Genet.* **24**, 221–230 (2008).
22. Tepass, U., Tanentzapf, G., Ward, R. & Fehon, R. Epithelial cell polarity and cell junctions in *Drosophila*. *Annu. Rev. Genet.* **35**, 747–784 (2001).
23. Bilder, D., Li, M. & Perrimon, N. Cooperative regulation of cell polarity and growth by *Drosophila* tumor suppressors. *Science* **289**, 113–116 (2000).
24. Bilder, D. & Perrimon, N. Localization of apical epithelial determinants by the basolateral PDZ protein Scribble. *Nature* **403**, 676–680 (2000).
25. Royer, C. & Lu, X. Epithelial cell polarity: a major gatekeeper against cancer? *Cell Death Differ.* **18**, 1470–1477 (2011).
26. Speck, O., Hughes, S. C., Noren, N. K., Kulikauskas, R. M. & Fehon, R. G. Moesin functions antagonistically to the Rho pathway to maintain epithelial integrity. *Nature* **421**, 83–87 (2003).
27. Thiery, J. P., Acloque, H., Huang, R. Y. J. & Nieto, M. A. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**, 871–890 (2009).
28. Pagliarini, R. A. & Xu, T. A genetic screen in *Drosophila* for metastatic behavior. *Science* **302**, 1227–1231 (2003).
29. Humbert, P. O. et al. Control of tumorigenesis by the Scribble/Dlg/Lgl polarity module. *Oncogene* **27**, 6888–6907 (2008).
30. Leung, C. T. & Brugge, J. S. Outgrowth of single oncogene-expressing cells from suppressive epithelial environments. *Nature* **482**, 410–413 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Bilder, M. Miura, Y. Yamashita, the Bloomington Stock Center, Vienna *Drosophila* RNAi Center, TRIP, and FlyTrap project for fly stocks and C. Doe, F. Matsuzaki and the Developmental Studies Hybridoma Bank for antibodies. We thank R. Fehon for suggestions. We thank T. Akiyama, A. Fritz, A. Ikmi and M. Szuperak for comments on the manuscript and L. Liang for technical advice. We also thank M. Gogol for graphical assistance, F. Guo for TEM support and L. Gutchesky for administrative support. This work was supported by the Stowers Institute for Medical Research and the Burroughs Wellcome Fund for Biomedical Research.

Author Contributions Y.-I.N. and M.C.G. conceived the project, designed the experiments and wrote the manuscript. Y.-I.N. and E.J.M. performed the experiments and analysed the data. A.K. and S.A.M. constructed the SPIM system and performed live imaging.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.C.G. (MG2@stowers.org).

METHODS

Fly stocks and genetics. The following stocks were used: *OreR* for wild-type control, A9-GAL4, UAS-*cnn*-GFP³¹, *nub*-GAL4, His2Av-mRFP (monomeric red fluorescent protein), UAS-Moe-GFP³², UAS-p35, UAS-reaper, UAS-GFP, UAS-Dicer-2, Asl-YFP³³, Nrg-GFP³⁴, scrib-GFP³⁵, hs-flp UAS-mCD8-GFP; *Tub*-GAL4 FRT82B *Tub*-GAL80 for MARCM³⁶ clones, *yw*; FRT82B, *yw*; FRT82B scrib^{1/}TM6C²³, hs-flp; *Act>y⁺>GAL4* UAS-GFP for FLP/FRT-mediated (FLP-out) clones, UAS-mud^{RNAi} (Bloomington 28074, Bloomington 35044), UAS-rok^{RNAi} (Bloomington 28797, Bloomington 34324), UAS-Moe^{RNAi} (VDRC 37917, Bloomington 33936), UAS-scrib^{RNAi} (Bloomington 35748), UAS-dlg^{RNAi} (Bloomington 33629, Bloomington 25780) for RNAi knockdown experiments (VDRC³⁷ and TRIP³⁸). Larvae were collected at 25 °C. Clones of RNAi constructs or mutants were generated with a 1-h heat shock at 48 h after egg lay.

Immunofluorescence. The following antibodies and dyes were used: rabbit anti-phospho-histone H3 (1:1,000, Millipore), mouse anti-phospho-histone H3 (1:2,000, Millipore), mouse anti- γ -tubulin (1:1,000, Sigma), rat anti- α -tubulin (1:250, Serotec), rabbit anti-phospho-MOE (1:100, Cell Signaling), rabbit anti-cleaved CASP3 (1:500, Cell Signaling), guinea pig anti-CNN (1:1,000, F. Matsuzaki), mouse anti-MUD (1:50, F. Matsuzaki), rabbit anti-SCRIB (1:5,000, C. Doe), rabbit anti-BAZ (1:500, F. Matsuzaki), mouse anti-DLG (1:500, DSHB), mouse anti-ARM (1:200, DSHB), rat anti-ECAD (1:25, DSHB), mouse anti-CORA (1:100, C615.16, DSHB), mouse anti-FAS3 (1:100, DSHB), mouse anti-MMP1 (1:100, 1:1:1 cocktail of 5H7B11, 3A6B4, 3B8D12, DSHB), rabbit anti-aPKC- ζ C20 (1:1,000, Santa Cruz Biotechnology), fluorescent second antibodies (1:500, Invitrogen), Alexa Phalloidin 488, 546 (1:500, Invitrogen) and Hoechst 33342 (2 μ M, Thermo Scientific).

Drug treatments. CytoD (5 μ M, Calbiochem) and Y-27632 (ref. 39) (1 mM, Calbiochem) were diluted in PBS and wing discs were incubated for 30 min as described⁷. These concentrations do not disrupt wing epithelial morphology during the assay period.

Imaging and image analysis. Confocal images were collected with a \times 63 glycerol or \times 40 oil objective lens on a Leica SP5 AOBS confocal microscope system (Leica Microsystems). To measure mitotic spindle orientation, images of mitotic spindles were taken by optical cross section (xz scan). The angle between the line perpendicular to the epithelium and the line through the spindle poles was measured with Fiji/ImageJ (Supplementary Fig. 1d). On the basis of the maximal outer limits defined by the average length of the septate junctions ($25.6 \pm 1.3^\circ$, $n = 91$), we

scored spindles as abnormal when angular deviation from planar (θ), $\theta \geq 30^\circ$. Values $60 \leq \theta \leq 90^\circ$ were considered strong misorientation and values $30 \leq \theta \leq 60^\circ$ were considered mild. In Figs 1–3, prometaphase/metaphase cells are shown for the quantification of mitotic spindle orientation. Spindle orientation in anaphase/telophase cells are in Supplementary Information.

SPIM. Wing discs were embedded in 1% agarose in *Drosophila* Ringers' solution. Embedded wing discs were then placed in a chamber filled with fly media⁴⁰ and imaged every minute for 2–3 h at approximately 21 °C using SPIM⁴¹. Fifty Z-sections with 1- μ m spacing were collected with an exposure time of 27 ms using 488 nm and 561 nm laser excitation scanned sequentially and detected using a W Plan-Apochromat \times 40/1.0 NA objective. Further details can be found in Supplementary Information.

31. Megrav, T. L., Kilaru, S., Turner, F. R. & Kaufman, T. C. The centrosome is a dynamic structure that ejects PCM flares. *J. Cell Sci.* **115**, 4707–4718 (2002).
32. Bloor, J. W. & Kiehart, D. P. zipper nonmuscle myosin-II functions downstream of PS2 integrin in *Drosophila* myogenesis and is necessary for myofibril formation. *Dev. Biol.* **239**, 215–228 (2001).
33. Varmark, H. et al. Asterless is a centriolar protein required for centrosome function and embryo development in *Drosophila*. *Curr. Biol.* **17**, 1735–1745 (2007).
34. Morin, X., Daneman, R., Zavortink, M. & Chia, W. A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc. Natl Acad. Sci. USA* **98**, 15050–15055 (2001).
35. Buszczak, M. et al. The carnegie protein trap library: a versatile tool for *Drosophila* developmental studies. *Genetics* **175**, 1505–1531 (2007).
36. Lee, T. & Luo, L. Mosaic analysis with a repressible cell marker for studies of gene function in neuronal morphogenesis. *Neuron* **22**, 451–461 (1999).
37. Dietzl, G. et al. A genome-wide transgenic RNAi library for conditional gene inactivation in *Drosophila*. *Nature* **448**, 151–156 (2007).
38. Ni, J. Q. et al. A genome-scale shRNA resource for transgenic RNAi in *Drosophila*. *Nature Methods* **8**, 405–407 (2011).
39. Uehata, M. et al. Calcium sensitization of smooth muscle mediated by a Rho-associated protein kinase in hypertension. *Nature* **389**, 990–994 (1997).
40. Gibson, W. T. et al. Control of the mitotic cleavage plane by local epithelial topology. *Cell* **144**, 427–438 (2011).
41. Keller, P. J., Schmidt, A. D., Wittbrodt, J. & Stelzer, E. H. K. Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. *Science* **322**, 1065–1069 (2008).

Co-crystal structure of a T-box riboswitch stem I domain in complex with its cognate tRNA

Jinwei Zhang¹ & Adrian R. Ferré-D'Amare¹

In Gram-positive bacteria, T-box riboswitches regulate the expression of aminoacyl-tRNA synthetases and other proteins in response to fluctuating transfer RNA aminoacylation levels under various nutritional states¹. T-boxes reside in the 5'-untranslated regions of the messenger RNAs they regulate, and consist of two conserved domains. Stem I contains the specifier trinucleotide that base pairs with the anticodon of cognate tRNA. 3' to stem I is the antiterminator domain, which base pairs with the tRNA acceptor end and evaluates its aminoacylation state². Despite high phylogenetic conservation and widespread occurrence in pathogens, the structural basis of tRNA recognition^{3,4} by this riboswitch remains ill defined. Here we demonstrate that the ~100-nucleotide T-box stem I is necessary and sufficient for specific, high-affinity (dissociation constant (K_d) ~150 nM) tRNA binding, and report the structure of *Oceanobacillus iheyensis* glyQ stem I in complex with its cognate tRNA at 3.2 Å resolution. Stem I recognizes the overall architecture of tRNA in addition to its anticodon, something accomplished by large ribonucleoproteins such as the ribosome, or proteins such as aminoacyl-tRNA synthetases⁵, but is unprecedented for a compact mRNA domain. The C-shaped stem I cradles the L-shaped tRNA, forming an extended (1,604 Å²) intermolecular interface. In addition to the specifier–anticodon interaction, two interdigitated T-loops near the apex of stem I stack on the tRNA elbow in a manner analogous to those of the J11/12–J12/11 motif⁶ of RNase P and the L1 stalk⁷ of the ribosomal E-site. Because these ribonucleoproteins and T-boxes are unrelated, this strategy to recognize a universal tRNA feature probably evolved convergently. Mutually induced fit of stem I and the tRNA exploiting the intrinsic flexibility of tRNA and its conserved post-transcriptional modifications results in high shape complementarity, which in addition to providing specificity and affinity, globally organizes the T-box to orchestrate tRNA-dependent transcription regulation.

T-box riboswitches contain conserved stem I and antiterminator domains separated by a variable linker. In some T-boxes, this linker contains three stem-loops. In the tRNA^{Gly}-specific glyQ and glyQS T-boxes, there is only one short stem-loop and a ~13-nucleotide single-stranded segment (Extended Data Fig. 1). We analysed the contribution of different T-box segments to tRNA binding by isothermal titration calorimetry (ITC). Full-length *Bacillus subtilis* glyQS T-box, a 3' truncation leaving only the antiterminator residues that base pair with the tRNA acceptor end, and an isolated stem I domain all bind tRNA^{Gly} with comparable affinity (Fig. 1a–c). Therefore, stem I provides essentially all the binding energy to tRNA. Proximal or distal truncations of stem I abrogate binding (Extended Data Fig. 2). Mutagenesis shows that high-affinity binding is contingent on cognate specifier–anticodon base pairing (Extended Data Table 1). However, stem I recognizes tRNA elements outside of the anticodon stem-loop (ASL), because an isolated ASL derived from tRNA^{Gly} does not bind appreciably (Extended Data Fig. 2). We conclude that stem I is necessary and sufficient for specific, high-affinity tRNA binding by the T-box riboswitch.

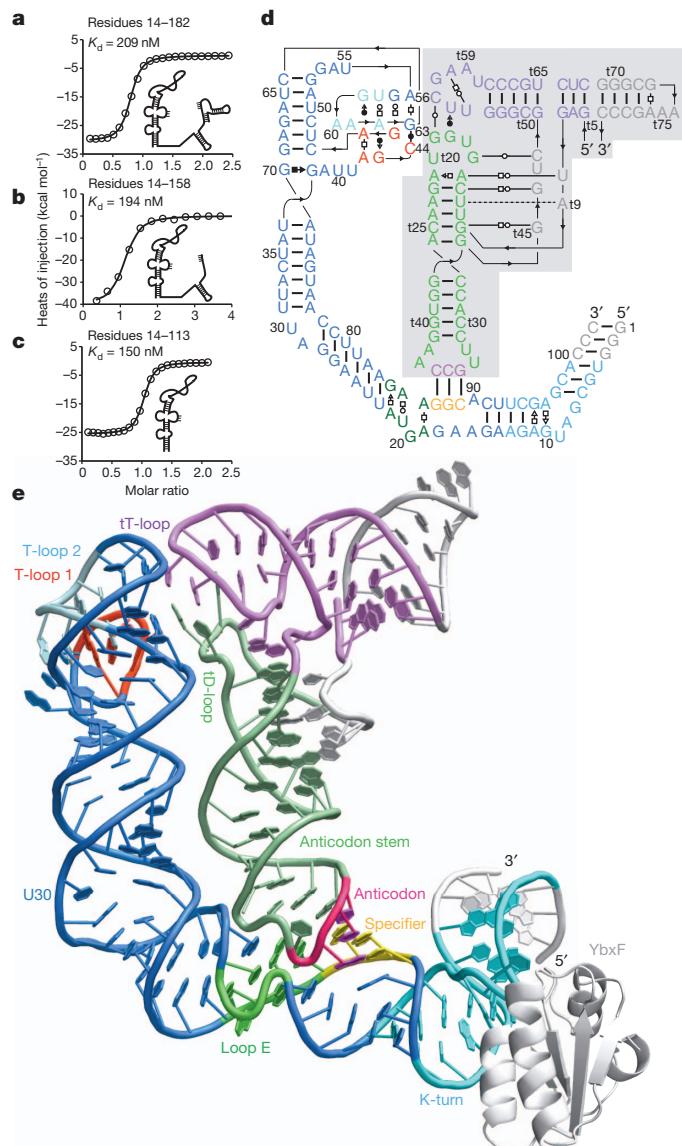


Figure 1 | Overall structure of the T-box stem I in complex with tRNA. **a**, ITC analysis of tRNA binding by full-length glyQS T-box (residues 14–182). **b**, Binding by a 3'-truncated T-box (residues 14–158). **c**, Binding by an isolated T-box stem I (residues 14–113). **d**, Sequence and secondary structure of the co-crystallized glyQ stem I and tRNA^{Gly} RNAs. Leontis–Westhof³⁰ symbols denote non-canonical base pairs. Lines with embedded arrowheads denote chain connectivity. The tRNA (shaded) is numbered conventionally ('t' precedes tRNA residues). **e**, Cartoon of the complex structure. Colour-coding as in **d**; segments altered to facilitate crystallization and the YbxF protein are in white.

¹National Heart, Lung and Blood Institute, 50 South Drive, MSC 8012, Bethesda, Maryland 20892-8012, USA.

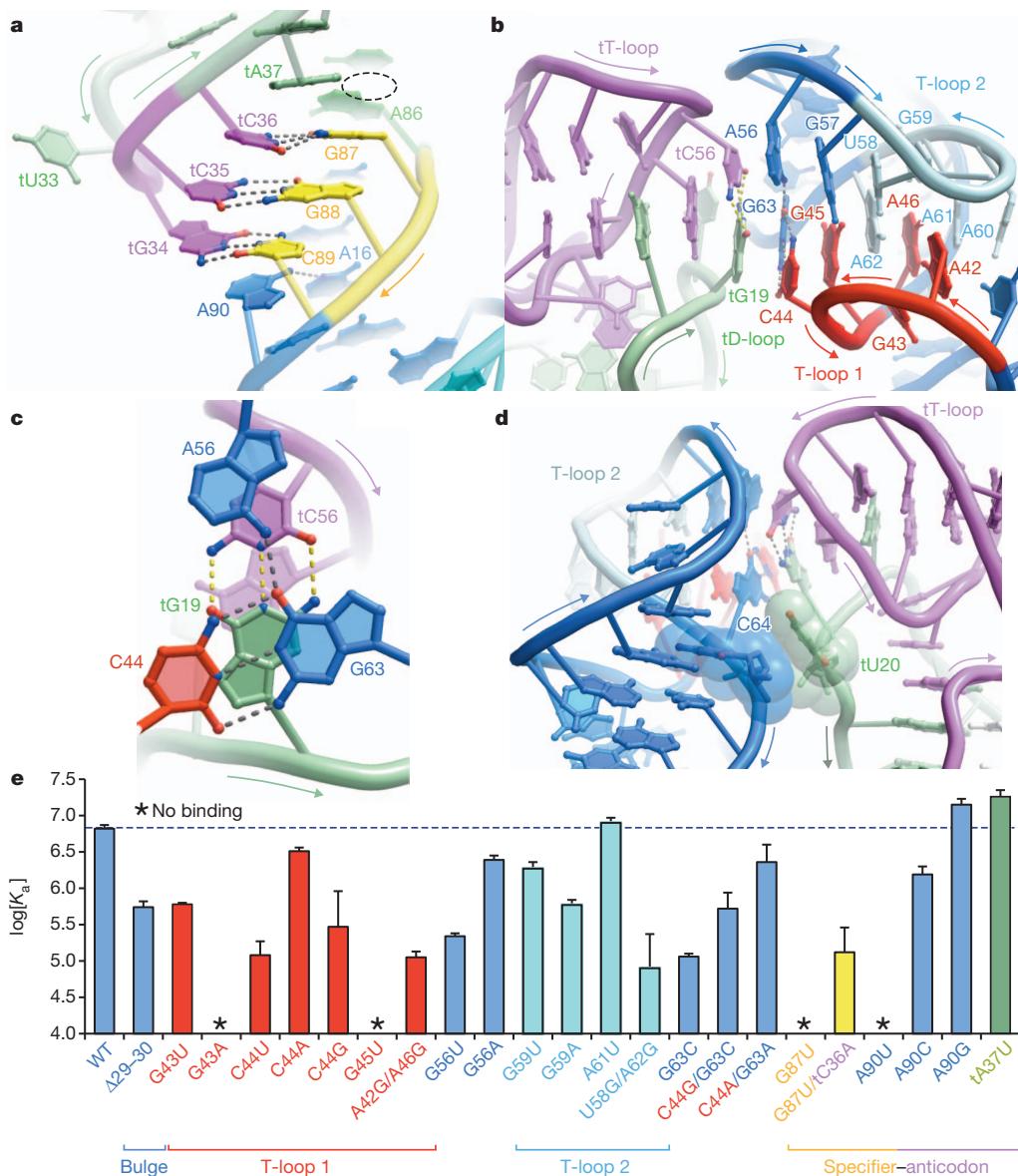


Figure 2 | Interactions between stem I and tRNA. **a**, Specifier–anticodon interaction. Dashed lines denote hydrogen bonds. Displacement of A86 forms a pocket (dashed oval). **b**, The two interdigitated T-loops at the distal end of stem I stack on the tRNA elbow. **c**, Stacking of the distal stem I base triple on the

tRNA elbow. Yellow dashed lines denote tRNA elbow base pairing. **d**, tRNA residue tU20 flips out to stack with the stem I C64 ribose (van der Waals surfaces of interacting residues shown). **e**, Mutagenesis and ITC analysis of selected stem-I-tRNA interactions. Data are mean and s.e.m. ($n \geq 2$).

To understand how an RNA domain of only ~ 100 residues can recognize a 76-nucleotide tRNA, we co-crystallized stem I from the *glyQ* T-box riboswitch of *Oceanobacillus iheyensis* (Fig. 1d) with tRNA^{Gly} and the bacterial K-turn binding protein YbxF, and solved its structure at 3.2 \AA resolution (Methods). Stem I folds into an irregular helix of $\sim 110\text{ \AA}$ contour length that cradles tRNA (Fig. 1e). Several RNA structural motifs are arranged linearly in stem I, consistent with phylogenetic conservation^{1,8–11}. The proximal end of stem I contains a K-turn¹², which introduces a 120° bend into the helical path. YbxF recognizes the K-turn as seen previously in its SAM-I riboswitch complex¹³. Immediately distal is the specifier trinucleotide, which base pairs with the tRNA anticodon, as predicted genetically². This is followed by a loop E motif¹⁴, with the extruded G20 forming an S-turn. The helical trajectory bends $\sim 65^\circ$ at a dinucleotide bulge (residues 29–30), allowing stem I to track the tRNA. A bulge (but not its sequence) is conserved at this location in T-box stem I domains¹¹ (Extended Data Figs 1 and 3). The phylogenetically conserved secondary structure of the distal portion of stem I contains another bulge followed by an apical loop (Extended Data Fig. 1). In our co-crystal

structure, this bulge and the loop each folds into a pentanucleotide T-loop¹⁵. The two T-loops interdigitate and recognize the structurally conserved elbow of tRNA. Association with tRNA^{Gly} buries $1,604\text{ \AA}^2$ of the solvent-accessible surface area of stem I (Extended Data Fig. 3). Of these, 939 \AA^2 are at the specifier–anticodon interaction and three adjacent interfaces where the backbones of loop E and two stem I segments on the same helical face contact backbone atoms of the tRNA anticodon stem. The remaining 665 \AA^2 are at the apical region of stem I, which contacts two D-loop residues and one T-loop residue of the tRNA. The solvent-accessible area of tRNA buried by the T-box is comparable to those buried by several aminoacyl-tRNA synthetases (ARSs)⁵, but much larger than the ligand surface areas buried by most riboswitches¹⁶.

T-boxes have a conserved purine¹¹ immediately 3' to the three specifier nucleotides. Our structure reveals that this nucleotide (A90) stacks underneath the wobble base pair (C89•tG34; tRNA residue numbers are preceded by 't') and also hydrogen bonds to A16 (Fig. 2a). A similar stabilizing interaction occurs in the P-site of the ribosome, where the universally conserved C1400 of 16S ribosomal

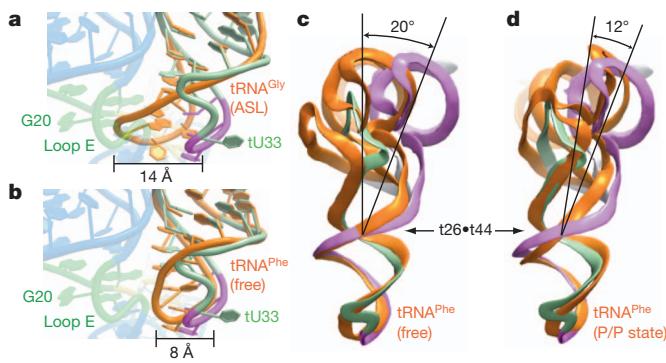


Figure 3 | Induced fit of tRNA by stem I binding. **a**, Superposition of ASL of free tRNA^{Gly} (Protein Data Bank (PDB) accession code 2LB1; orange) with the co-crystal structure. Steric clash between the free structure and the loop E motif (green) is evident. **b**, Superposition of the ASL of free tRNA^{Phe} (PDB code 1EHZ, orange) with the co-crystal structure. Note extrusion of tU33. **c**, Comparison of stem-I-bound tRNA^{Gly} with free tRNA^{Phe} (orange) seen from the direction of the elbow. The approximate location of the hinge at the t26•t44 pair is indicated. **d**, Comparison of stem-I-bound tRNA^{Gly} with the ribosome-bound P/P state tRNA (PDB code 4GD2, orange).

tRNA stacks underneath the tRNA–mRNA wobble pair^{7,17}. Near-symmetrically to A90, tA37 stacks on top of the specifier–anticodon duplex. A purine typically occupies tRNA position t37 and is often post-transcriptionally modified into a larger nucleobase. The nucleotide preceding the first specifier nucleotide of stem I (A86) is unstacked from the specifier–anticodon duplex because of its participation in the sheared A•A pair that initiates the loop E motif of stem I. This unstacking produces a pocket that could accommodate large, modified t37 nucleobases (Fig. 2a and Extended Data Fig. 4).

The two interdigitated T-loops and adjacent nucleotides from the distal end of stem I together form a compact structural module consisting of six stacked layers (Extended Data Fig. 5). The most distal of these is formed by three co-planar bases, C44, A56 and G63. In the stem-I–tRNA complex, this base triple stacks on the conserved tRNA base pair tG19•tC56 (Fig. 2b). This conserved tertiary pair joins the apices of the tRNA D- and T-loops and forms the ‘elbow’ of all elongator tRNAs. Thus, the two T-loops of stem I stack on the T- and D-loops of tRNA using the flat molecular surfaces formed by two interlocked loops in each RNA (Fig. 2b, c). The nucleobase of tU20, another conserved residue from the D-loop of tRNAs, extends the interface by stacking on the ribose of stem I residue C64 (Fig. 2d)—itself part of a conserved base pair¹¹. Similar extrahelical stacking of t20 has been observed in tRNA–ARS complexes^{18,19}. Position t20 is post-transcriptionally modified in most tRNAs to dihydrouridine, which stabilizes the C2'-endo conformation of the nucleotide²⁰, potentially enhancing stacking with the ribose of C64 in stem I. The functional importance of several stem-I–tRNA interactions revealed by our co-crystal analysis is supported by mutagenesis and ITC analysis (Fig. 2e). In particular, alterations to the specifier, the adjacent purine (A90), the core and the distal surface of the interdigitated T-loops are all strongly detrimental.

Sequence analysis and modelling^{9,10,21} have previously suggested that the two interdigitated T-loops of stem I fold and function similarly to the J11/12–J12/11 internal loops of RNase P (ref. 22), and the L1 stalk in the E-site of the 50S ribosomal subunit²³. Our co-crystal structure confirms this. However, the stem I T-loops bind to the tRNA elbow in the opposite orientation from those of RNase P and the ribosome^{6,7} (Extended Data Fig. 6). Because there is no evidence that 23S rRNA, RNase P and T-boxes share an ancestor, the use of interdigitated T-loops to recognize the tRNA elbow has probably evolved independently at least three times.

Association with stem I distorts tRNA locally and globally. The tRNA^{Gly} ASL rearranges after T-box binding, probably to avoid steric

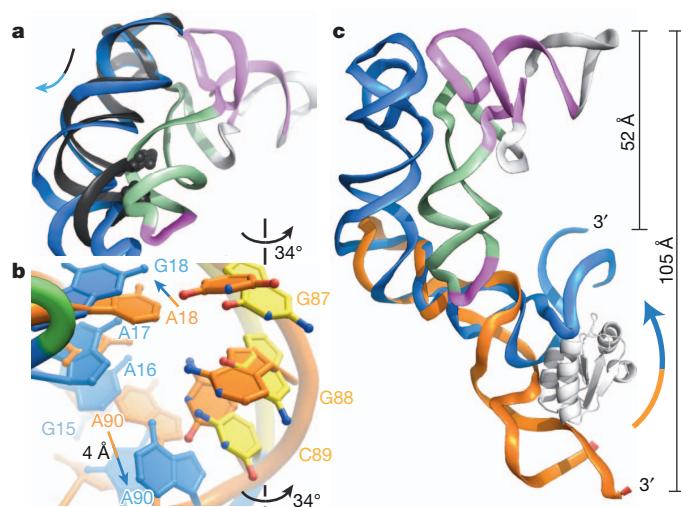


Figure 4 | Stem I reorganization by tRNA binding. **a**, Superposition of an isolated stem I distal fragment (PDB code 4JRC, black) and the complex. Clash between the helix (partly shown as spheres) of the fragment and the tRNA is relieved by flexing below the T-loops in the complex. **b**, Superposition of the specifier free (PDB code 2KZL, orange) and bound to tRNA. Specifier nucleotides rotate outward, and A90 is displaced. **c**, Superposition of an isolated stem I proximal fragment (PDB code 2KZL, orange) and the complex. tRNA binding induces bending of stem I (arrow), bringing its 3' terminus ~50 Å closer to the acceptor end of tRNA.

clash. Its open structure free in solution²⁴ superimposes poorly on that in the complex (Fig. 3a, b). Comparison with ribosome structures reveals that, similar to tRNAs engaged in translation^{7,25–27}, the stem-I-bound tRNA^{Gly} bends near the t26•t44 pair at the junction of the anticodon and D stems. This bend resembles that of a P/P state tRNA (~10° relative to the free tRNA^{Phe}), but is more pronounced (~20°; Fig. 3c, d). Thus, in addition to deforming the anticodon loop, the T-box seems to exploit the intrinsically flexible t26•t44 hinge to shift the tRNA elbow by ~11 Å, bringing it into contact with the distal base triple of stem I (Extended Data Fig. 7).

Stem I also undergoes pronounced conformational changes after tRNA binding. A structure²¹ of an apical fragment of stem I shows good agreement of the interdigitated T-loops; however, a ~20° hinge motion proximal to the T-loops must occur concomitant with binding to avoid steric clash with the tRNA (Fig. 4a). Comparison with the solution structure²⁸ of a proximal fragment of a *tyrS* T-box shows that specifier residues rotate on average 34° out of the minor groove to base pair with the anticodon. This rotation propagates, moving the functionally crucial A90 by 4 Å so it stacks under the specifier–anticodon duplex (Figs 2e and 4b). The propagation also results in continuous stacking of the phylogenetically conserved purines²⁹ (A8–A19) on the 5' strand of stem I. Markedly different from the extended structure²⁸ exhibited in solution in the absence of tRNA, the proximal residues of stem I fold into a canonical K-turn in the tRNA complex. This results in a marked contraction of the distance separating the 3'-terminus of stem I from the acceptor end of tRNA (Fig. 4c). This distance needs to be bridged by the short linker separating the glyQS stem I from the antiterminator (Extended Data Fig. 1) for the T-box riboswitch to function. A role for the K-turn in organizing the large-scale architecture of the T-box–tRNA complex reconciles its reported requirement *in vivo*⁸ and its lack of contribution to tRNA binding affinity *in vitro* (Extended Data Fig. 2 and Extended Data Table 1). Therefore, the mutually induced fit of stem I and tRNA not only gives rise to specific, high-affinity binding, but also organizes the overall architecture of the 5'-untranslated region to orchestrate the aminoacylation-dependent transcriptional termination decision.

Despite having different sequences and being extensively modified post-transcriptionally, all elongator tRNAs share a common overall

architecture as well as a flexible hinge that allows efficient transit through the ribosome. Our T-box-tRNA co-crystal structure reveals how a gene-regulatory mRNA domain exploits all these molecular features of tRNA in addition to the anticodon sequence for specific recognition. Phylogenetic conservation indicates that T-boxes use a conserved structural scaffold to recognize divergent tRNAs. Our analysis therefore provides a framework with which to understand the coevolution of interacting RNA molecules.

METHODS SUMMARY

O. iheyensis glyT Box stem I was modified for crystallization by deleting one base pair from the distal side of the K-turn (above the A11•G95 pair) and introducing three base pairs at its proximal end. *O. iheyensis* tRNA^{Gly}_{GCC} was circularly permuted and its acceptor stem was capped with a GAAA tetraloop (Fig. 1d). The two RNAs and one equivalent of selenomethionyl *B. subtilis* YbfX (200 μM complex in a buffer consisting of 50 mM HEPES-KOH, pH 7.0, 100 mM KCl, 20 mM MgCl₂ and 5 mM tris(2-carboxyethyl)phosphine (TCEP)) were co-crystallized at 21 °C by vapour diffusion by mixing with an equal volume of a reservoir consisting of 50 mM Bis-Tris (HCl), pH 6.5, 300 mM Li₂SO₄ and 20% (w/v) PEG3350, supplemented with 0.2% (w/v) low-melting-point agarose. Co-crystals with the symmetry of space group C222₁ were incubated in a harvesting solution consisting of 50 mM Bis-Tris (HCl), pH 6.5, 100 mM KCl, 20–50 mM SrCl₂, 40–45% (w/v) PEG3350 and 5 mM TCEP for 16 h before flash-freezing in liquid nitrogen. Diffraction data were collected at 100 K at beamlines 5.0.1 and 5.0.2 of the Advanced Light Source (ALS), and beamlines 24-ID-C and 24-ID-E of the Advanced Photon Source (APS). Single wavelength anomalous dispersion (SAD) phases from selenomethionine and Sr²⁺ ions were combined with molecular replacement phases obtained using tRNA and a K-turn-YbfX complex as search models to produce initial electron density maps. The structure, which contains two stem-I-tRNA^{Gly}-YbfX ternary complexes per crystallographic asymmetric unit, has been refined to $R_{\text{free}} = 25.2\%$, and has a mean coordinate precision of 0.47 Å (Extended Data Table 2). The two ternary complexes in the asymmetric unit superimpose closely. Biochemical and ITC analysis of the T-box-tRNA interaction was performed using RNAs derived from the *B. subtilis* glyQS sequence (Extended Data Fig. 1).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 13 May; accepted 9 July 2013.

Published online 28 July 2013.

- Green, N. J., Grundy, F. J. & Henkin, T. M. The T box mechanism: tRNA as a regulatory molecule. *FEBS Lett.* **584**, 318–324 (2010).
- Grundy, F. J. & Henkin, T. M. tRNA as a positive regulator of transcription antitermination in *B. subtilis*. *Cell* **74**, 475–482 (1993).
- Yousef, M. R., Grundy, F. J. & Henkin, T. M. Structural transitions induced by the interaction between tRNA^{Gly} and the *Bacillus subtilis* glyQS T box leader RNA. *J. Mol. Biol.* **349**, 273–287 (2005).
- Yousef, M. R., Grundy, F. J. & Henkin, T. M. tRNA requirements for glyQS antitermination: a new twist on tRNA. *RNA* **9**, 1148–1156 (2003).
- Perona, J. J. & Hadd, A. Structural diversity and protein engineering of the aminoacyl-tRNA synthetases. *Biochemistry* **51**, 8705–8729 (2012).
- Reiter, N. J. et al. Structure of a bacterial ribonuclease P holoenzyme in complex with tRNA. *Nature* **468**, 784–789 (2010).
- Korostelev, A., Trakhanov, S., Laurberg, M. & Noller, H. Crystal structure of a 70S ribosome-tRNA complex reveals functional interactions and rearrangements. *Cell* **126**, 1065–1077 (2006).
- Winkler, W. C., Grundy, F. J., Murphy, B. A. & Henkin, T. M. The GA motif: an RNA element common to bacterial antitermination systems, rRNA, and eukaryotic RNAs. *RNA* **7**, 1165–1172 (2001).
- Winkler, W. *RNA Elements Required for T Box Antitermination*. PhD thesis, Ohio State Univ. (2002).
- Lehmann, J., Jossinet, F. & Gautheret, D. A universal RNA structural motif docking the elbow of tRNA in the ribosome, RNase P and T-box leaders. *Nucleic Acids Res.* **41**, 5494–5502 (2013).

- Gutiérrez-Preciado, A., Henkin, T. M., Grundy, F. J., Yanofsky, C. & Merino, E. Biochemical features and functional implications of the RNA-based T-Box regulatory mechanism. *Microbiol. Mol. Biol. Rev.* **73**, 36–61 (2009).
- Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J.* **20**, 4214–4221 (2001).
- Baird, N. J., Zhang, J., Hamma, T. & Ferré-D'Amare, A. R. YbfX and YlxQ are bacterial homologs of L7Ae and bind K-turns but not K-loops. *RNA* **18**, 759–770 (2012).
- Leontis, N. B. & Westhof, E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134–1153 (1998).
- Chan, C. W., Chetnani, B. & Mondragón, A. Structure and function of the T-loop structural motif in noncoding RNAs. *Wiley Interdiscip Rev. RNA* <http://dx.doi.org/10.1002/wrna.1175> (23 June 2013).
- Edwards, T. E., Klein, D. J. & Ferré-D'Amare, A. R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr. Opin. Struct. Biol.* **17**, 273–279 (2007).
- Selmer, M. et al. Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* **313**, 1935–1942 (2006).
- Rock, F. L. et al. An antifungal agent inhibits an aminoacyl-tRNA synthetase by trapping tRNA in the editing elite. *Science* **316**, 1759–1761 (2007).
- Fukai, S. et al. Structural basis for double-sieve discrimination of L-valine from L-isoleucine and L-threonine by the complex of tRNA^{Val} and valyl-tRNA synthetase. *Cell* **103**, 793–803 (2000).
- Dalluge, J. J., Hashizume, T., Sopchik, A. E., McCloskey, J. A. & Davis, D. R. Conformational flexibility in RNA: the role of dihydrouridine. *Nucleic Acids Res.* **24**, 1073–1079 (1996).
- Grigg, J. C. et al. T box RNA decodes both the information content and geometry of tRNA to affect gene expression. *Proc. Natl. Acad. Sci. USA* **110**, 7240–7245 (2013).
- Krasilnikov, A. S., Yang, X., Pan, T. & Mondragón, A. Crystal structure of the specificity domain of ribonuclease P. *Nature* **421**, 760–764 (2003).
- Nikulin, A. et al. Structure of the L1 protuberance in the ribosome. *Nature Struct. Biol.* **10**, 104–108 (2003).
- Chang, A. T. & Nikonowicz, E. P. Solution nuclear magnetic resonance analyses of the anticodon arms of proteinogenic and nonproteinogenic tRNA^{Gly}. *Biochemistry* **51**, 3662–3674 (2012).
- Valle, M. et al. Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nature Struct. Biol.* **10**, 899–906 (2003).
- Dunkle, J. A. et al. Structures of the bacterial ribosome in classical and hybrid states of tRNA binding. *Science* **332**, 981–984 (2011).
- Schmeing, T. M. et al. The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* **326**, 688–694 (2009).
- Wang, J. & Nikonowicz, E. P. Solution structure of the K-turn and specifier loop domains from the *Bacillus subtilis* tyrS T-Box leader RNA. *J. Mol. Biol.* **408**, 99–117 (2011).
- Vitreschak, A. G., Mironov, A. A., Lyubetsky, V. A. & Gelfand, M. S. Comparative genomic analysis of T-box regulatory systems in bacteria. *RNA* **14**, 717–735 (2008).
- Leontis, N. B. & Westhof, E. Geometric nomenclature and classification of RNA base pairs. *RNA* **7**, 499–512 (2001).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the staff at beamlines 24-ID-C and 24-ID-E of APS and 5.0.1 and 5.0.2 of the ALS for crystallographic data collection support, G. Piszczeck for ITC support, R. Levine and D.-Y. Lee for help with mass spectrometry, K. Perry and K. R. Rajashankar for assistance with initial crystallographic phasing, and N. Baird, T. Hamma, C. Jones, M. Lau, J. Posakony, A. Roll-Mecak, O. Uhlenbeck and K. Warner for discussions. This work is partly based on research conducted at the APS on the Northeastern Collaborative Access Team (NE-CAT) beamlines, and at the ALS on the Berkeley Center for Structural Biology beamlines, which are both supported by the National Institute for General Medical Sciences, National Institutes of Health (NIH). Use of ALS and APS was supported by the US Department of Energy. This work was supported in part by the intramural program of the National Heart, Lung, and Blood Institute, NIH.

Author Contributions J.Z. and A.R.F.-D. designed experiments, J.Z. carried out all biochemistry and ITC experiments, J.Z. prepared co-crystals, J.Z. and A.R.F.-D. collected diffraction data, solved and refined the crystal structures, and wrote the manuscript.

Author Information Atomic coordinates and structure factor amplitudes for the T-box stem I in complex with tRNA^{Gly} and YbfX have been deposited at the Protein Data Bank under accession code 4LCK. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.R.F.-D. (adrian.ferre@nih.gov).

METHODS

RNA and protein preparation. Biochemical and ITC analysis of the T-box-tRNA interaction was performed using RNAs derived from the *B. subtilis* glyQS sequence (Extended Data Fig. 1). For crystallization, the *O. iheyensis* glyQ T-box stem I RNA was modified by deleting one base pair from the distal side of the K-turn (above the A11–G95 pair) and introducing three base pairs at its proximal end. The *O. iheyensis* tRNA^{Gly}_{GCC} was circularly permuted, and its acceptor stem was capped with a GAAA tetraloop (Fig. 1d). These alterations did not adversely affect the stem-I-tRNA association (Extended Data Fig. 2h and Extended Data Table 1). *B. subtilis* glyQS T-box and tRNA^{Gly} constructs, *O. iheyensis* glyQ T-box stem I RNA (A89C), and *O. iheyensis* tRNA^{Gly} (Fig. 1d) were transcribed *in vitro* using T7 RNA polymerase from double-stranded DNA templates generated by PCR, as described³¹, except that 2 ml of PCR reaction product was used for each 5-ml transcription. RNAs were purified by electrophoresis on 8% polyacrylamide, 8 M urea gels (29:1 acrylamide:bisacrylamide), electroeluted, washed once with 1 M KCl, desalted by ultrafiltration, and stored at 4 °C. Selenomethionyl *B. subtilis* YbfX was expressed and purified as described¹³.

ITC. To suppress oligomerization, tRNA^{Gly} was heated to 90 °C in water for 3 min and cooled to 4 °C over 2 min immediately before use. RNA samples for ITC experiments were equilibrated in a buffer consisting of 100 mM KCl, 50 mM HEPES-KOH, pH 7.0, and 10 mM MgCl₂ by ultrafiltration. ITC experiments were performed at least in duplicate at 20 °C with 20 μM T-box RNA in the cell and 200 μM tRNA or ASL in the syringe, using a MicroCal iTC₂₀₀ microcalorimeter (GE). Initial data analysis was performed as described^{32–34}; then, data from replicate experiments were fitted globally^{35,36} using SEDPHAT and NITPIC to arrive at the dissociation constants and uncertainties in Extended Data Table 1.

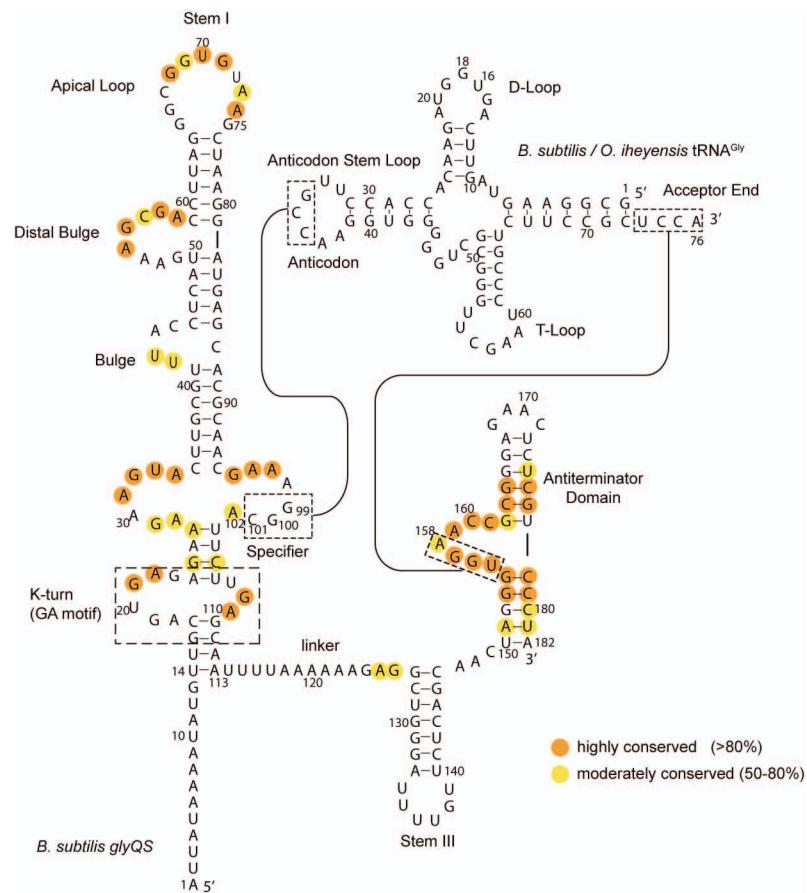
Co-crystallization and diffraction data collection. Engineered *O. iheyensis* tRNA^{Gly} (75 nucleotides) was heated to 90 °C in water for 3 min and cooled to 4 °C over 2 min, mixed with one equivalent of stem I RNA and incubated in the presence of 50 mM HEPES-KOH, pH 7.0, 100 mM KCl, 20 mM MgCl₂ and 5 mM tris(2-carboxyethyl)phosphine (TCEP) at 50 °C for 10 min and then at 37 °C for 30 min. One equivalent of selenomethionyl *B. subtilis* YbfX was then added. The solution was adjusted to 200 μM complex, 2 mM spermine, 0.2% (w/v) low-melting-point agarose and held at 37 °C. For crystallization, the complex was mixed 1:1 with a reservoir solution consisting of 50 mM Bis-Tris (HCl), pH 6.5, 300 mM Li₂SO₄ and 20% PEG3350. Plate-shaped crystals grew in 1–4 weeks to maximum dimensions of 300 × 300 × 50 μm³. After incubation in 50 mM Bis-Tris (HCl), pH 6.5, 100 mM KCl, 20–50 mM SrCl₂, 40–45% PEG3350 and 5 mM TCEP for 16 h, crystals were dissected out of the agarose using MicroSaws (Mitegen), mounted on 90° bent MicroLoops (Mitegen), and vitrified by plunging into liquid nitrogen. Co-crystals have the symmetry of space group C222₁. Unit cell dimensions are in Extended Data Table 2. Single-wavelength anomalous dispersion (SAD) data were collected at the selenium K edge (0.9792 Å) at 100 K at beamlines 24-ID-C and 24-ID-E of the Advanced Photon Source (APS), and indexed, integrated, and scaled using the NE-CAT RAPD pipeline that uses XDS³⁷ and Scala³⁸. Additional data were collected at beamlines 5.0.1 and 5.0.2 of the Advanced Light Source (ALS), and processed with HKL2000 (ref. 39). Data collection statistics are summarized in Extended Data Table 2.

Structure determination and refinement. Four selenium sites were identified from crystal I data using SHELX-D (ref. 40) through HKL2MAP (ref. 41). The resulting SAD phases were combined with model phases resulting from placing two truncated search models each of tRNA⁴² (PDB code 3L0U) and K-turn bound by YbfX¹³ (PDB code 3V7E) using MOLREP⁴³, in a strategy similar to that described⁴⁴. Subsequently, the composite ensemble (two copies each of tRNA and the K-turn–YbfX complex) was located in a merged, high-redundancy data set (crystals I–II–III) using PHASER⁴⁵, allowing completion of the anomalous atom substructure, which consists of eight selenium atom sites and three Sr²⁺ ions. SAD (mean overall figure of merit = 0.36) and model phases were combined using the molecular replacement–SAD pipeline implemented in PHENIX⁴⁶ and density-modified using RESOLVE⁴⁷, producing a substantially improved electron density map (Extended Data Fig. 8) with which manual model building could begin. Iterative rounds of manual model building⁴⁸, molecular replacement, and phase-combination produced a near-complete model. This model was then subjected to additional manual building interspersed with iterative rounds of rigid-body, simulated-annealing, and individual isotropic B-factor refinement against crystal I data using PHENIX. Refinement statistics are summarized in Extended

Data Table 2. The maximum likelihood coordinate precision of the current model is 0.47 Å. Ramachandran analysis shows that 92.4% of the amino acid residues are in the most favoured regions, 7.0% in allowed regions, and 0.6% (one residue) in disallowed regions. The two stem I and tRNA molecules in the asymmetric unit superimpose closely (root mean squared deviation (r.m.s.d.) of 1.1 and 0.7 Å, respectively). In the crystals, the flat surface formed by the interdigitated T-loops on the face opposite the tRNA elbow stacks on both the apical nucleobase of the GAAA tetraloop of the engineered tRNA, and the terminal (most proximal) base pair of stem I from symmetry-related molecules.

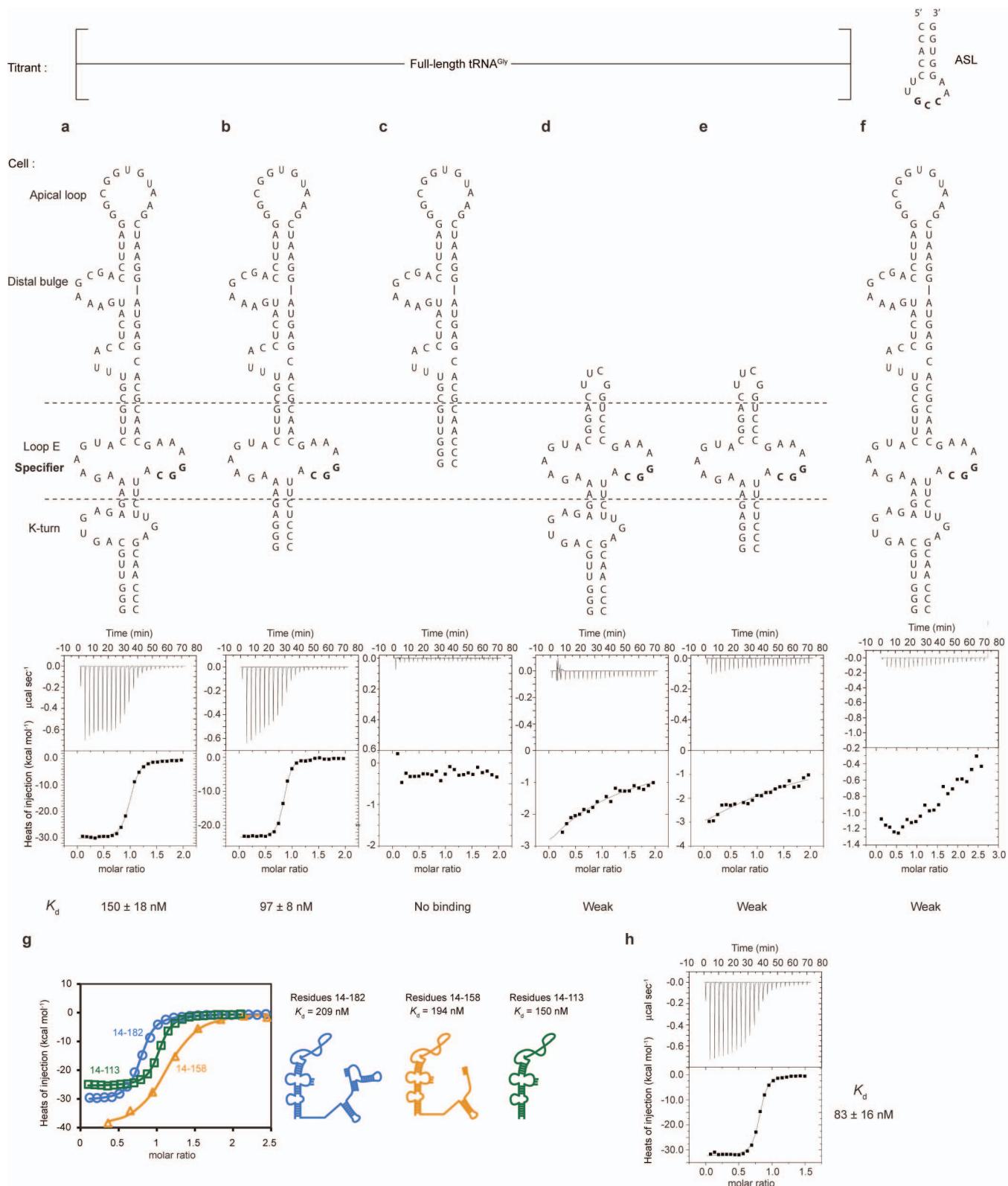
Comparisons were performed against the following structures: tRNA^{Gly} ASL (PDB code 2LBJ, ref. 24); tRNA^{Phe} (PDB code 1EHZ, ref. 49); the 70S ribosome with a P/P tRNA (PDB code 4GD2, ref. 26); a distal fragment of stem I (PDB code 4JRC, ref. 21); a proximal fragment of stem I (PDB ID: 2KZL, ref. 28); the loop E motif of 5S rRNA (PDB code 354D, ref. 50); RNase P holoenzyme bound to tRNA (PDB code 3Q1Q, ref. 6); the L1 stalk of the 50S ribosomal subunit (PDB code 1MZP, ref. 23); and the 70S ribosome with an E/E tRNA (PDB code 1VSA, ref. 7). The superposition shown in Fig. 4a used the interdigitated T-loops of the two structures. Those shown in Fig. 4b, c were generated by overlaying the loop E residues of the two structures. In both cases, r.m.s.d. was less than 1 Å. Solvent accessible surface areas were calculated with a probe radius of 1.4 Å. Structural figures were prepared with PyMol⁵¹, using chains A (YbfX), B (tRNA^{Gly}) and C (glyQ stem I).

31. Xiao, H., Edwards, T. E. & Ferré-D'Amaré, A. R. Structural basis for specific, high-affinity tetracycline binding by an *in vitro* evolved aptamer and artificial riboswitch. *Chem. Biol.* **15**, 1125–1137 (2008).
32. Baird, N. J. & Ferré-D'Amaré, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA* **16**, 598–609 (2010).
33. Kulshina, N., Edwards, T. E. & Ferré-D'Amaré, A. R. Thermodynamic analysis of ligand binding and ligand binding-induced tertiary structure formation by the thiamine pyrophosphate riboswitch. *RNA* **16**, 186–196 (2010).
34. Baird, N. J. & Ferré-D'Amaré, A. R. Modulation of quaternary structure and enhancement of ligand binding by the K-turn of tandem glycine riboswitches. *RNA* **19**, 167–176 (2013).
35. Schuck, P. Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophys. J.* **78**, 1606–1619 (2000).
36. Keller, S. *et al.* High-precision isothermal titration calorimetry with automated peak-shape analysis. *Anal. Chem.* **84**, 5066–5073 (2012).
37. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
38. Evans, P. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
39. Otwinowski, Z. & Minor, W. Processing of diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
40. Sheldrick, G. M. A short history of SHELX. *Acta Crystallogr. A* **64**, 112–122 (2008).
41. Pape, T. & Schneider, T. R. HKL2MAP: a graphical user interface for macromolecular phasing with SHELX programs. *J. Appl. Crystallogr.* **37**, 843–844 (2004).
42. Byrne, R. T., Konevega, A. L., Rodnina, M. V. & Antson, A. A. The crystal structure of unmodified tRNA^{Phe} from *Escherichia coli*. *Nucleic Acids Res.* **38**, 4154–4162 (2010).
43. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Crystallogr.* **30**, 1022–1025 (1997).
44. Xiao, H., Murakami, H., Suga, H. & Ferré-D'Amaré, A. R. Structural basis of specific tRNA aminoacylation by a small *in vitro* selected ribozyme. *Nature* **454**, 358–361 (2008).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
46. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D* **68**, 352–367 (2012).
47. Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
48. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
49. Shi, H. & Moore, P. B. The crystal structure of yeast phenylalanine tRNA at 1.93 Å resolution: a classic structure revisited. *RNA* **6**, 1091–1105 (2000).
50. Correll, C. C., Freeborn, B., Moore, P. B. & Steitz, T. A. Metals, motifs, and recognition in the crystal structure of a 5S rRNA domain. *Cell* **91**, 705–712 (1997).
51. DeLano, W. L. *The PyMOL Molecular Graphics System* (DeLano Scientific, 2002).
52. Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. Recurrent structural RNA motifs, isostericity matrices and sequence alignments. *Nucleic Acids Res.* **33**, 2395–2409 (2005).



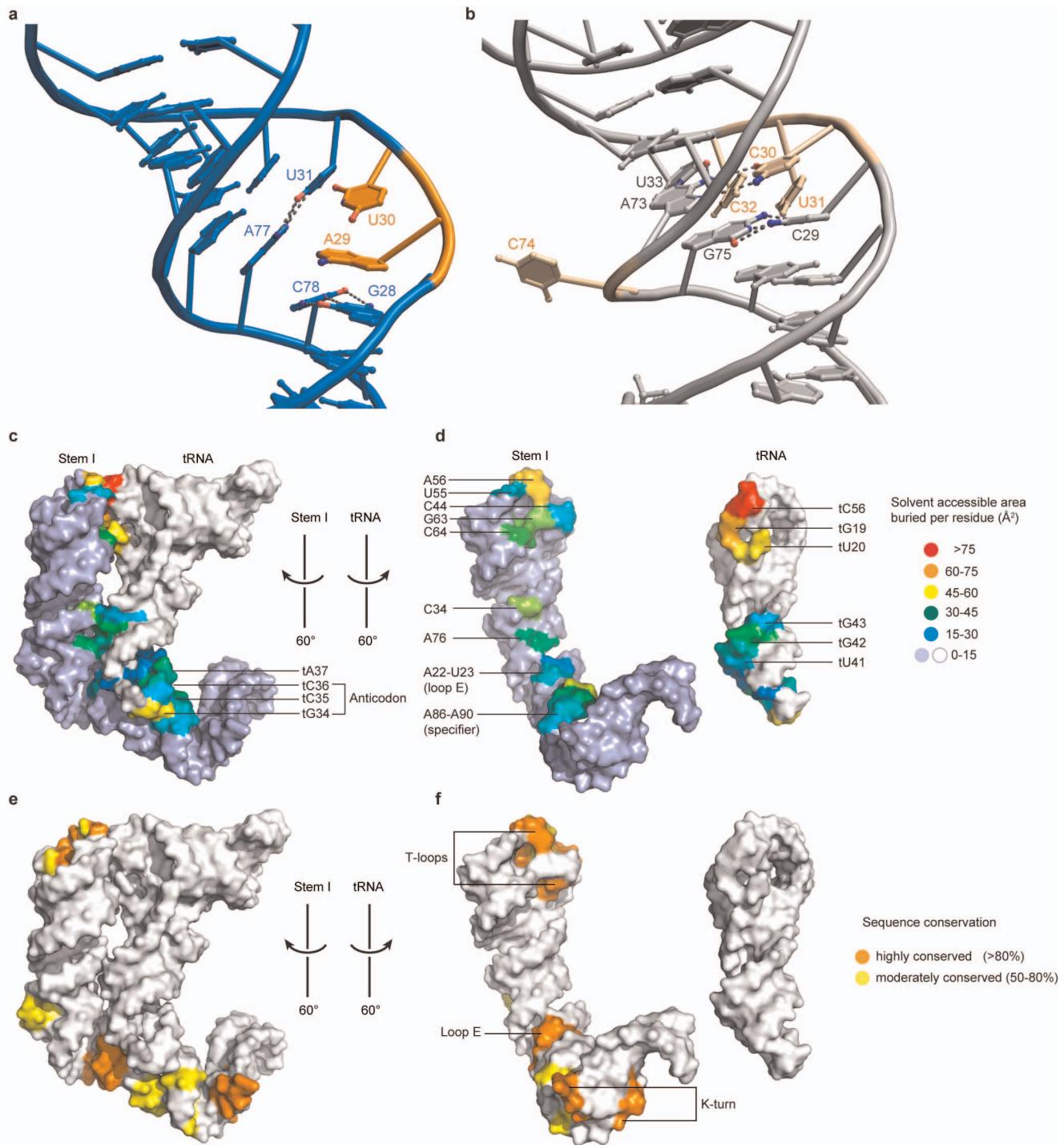
Extended Data Figure 1 | Sequence of the *B. subtilis* glyQS T-box. Unlike some other T-boxes, tRNA^{Gly}-responsive T-boxes (such as the *B. subtilis* glyQS T-box and the closely related *O. iheyensis* glyQ T-box) do not contain stems II and II A/B in the linker region. Conservation across known T-boxes is indicated in orange and yellow circles (from ref. 11). Base pairing between the specifier

sequence in stem I and the antiterminator domain of the T-box and the anticodon and acceptor end, respectively, of tRNA^{Gly} is indicated. The functionally important K-turn^a is boxed. The sequences of tRNA^{Gly} of *B. subtilis* and *O. iheyensis* are identical.



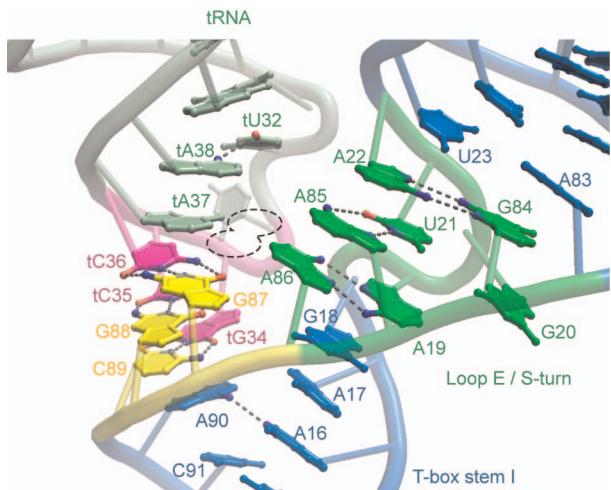
Extended Data Figure 2 | Representative isothermal calorimetric titrations. a–f, Full-length tRNA^{Gly} (a–e) and an isolated tRNA^{Gly} ASL (f) were titrated into various truncated *B. subtilis* glyQS T-box stem I constructs (Methods). The specifier trinucleotide GGC is in bold. g, Superposition of the experimental data

and fits for the experiments shown in Fig. 1a–c, colour-coded to correspond to schematic RNA construct depictions. h, ITC analysis of binding by the engineered tRNA^{Gly} construct used for co-crystallization. Results of all fits (Methods) are in Extended Data Table 1.

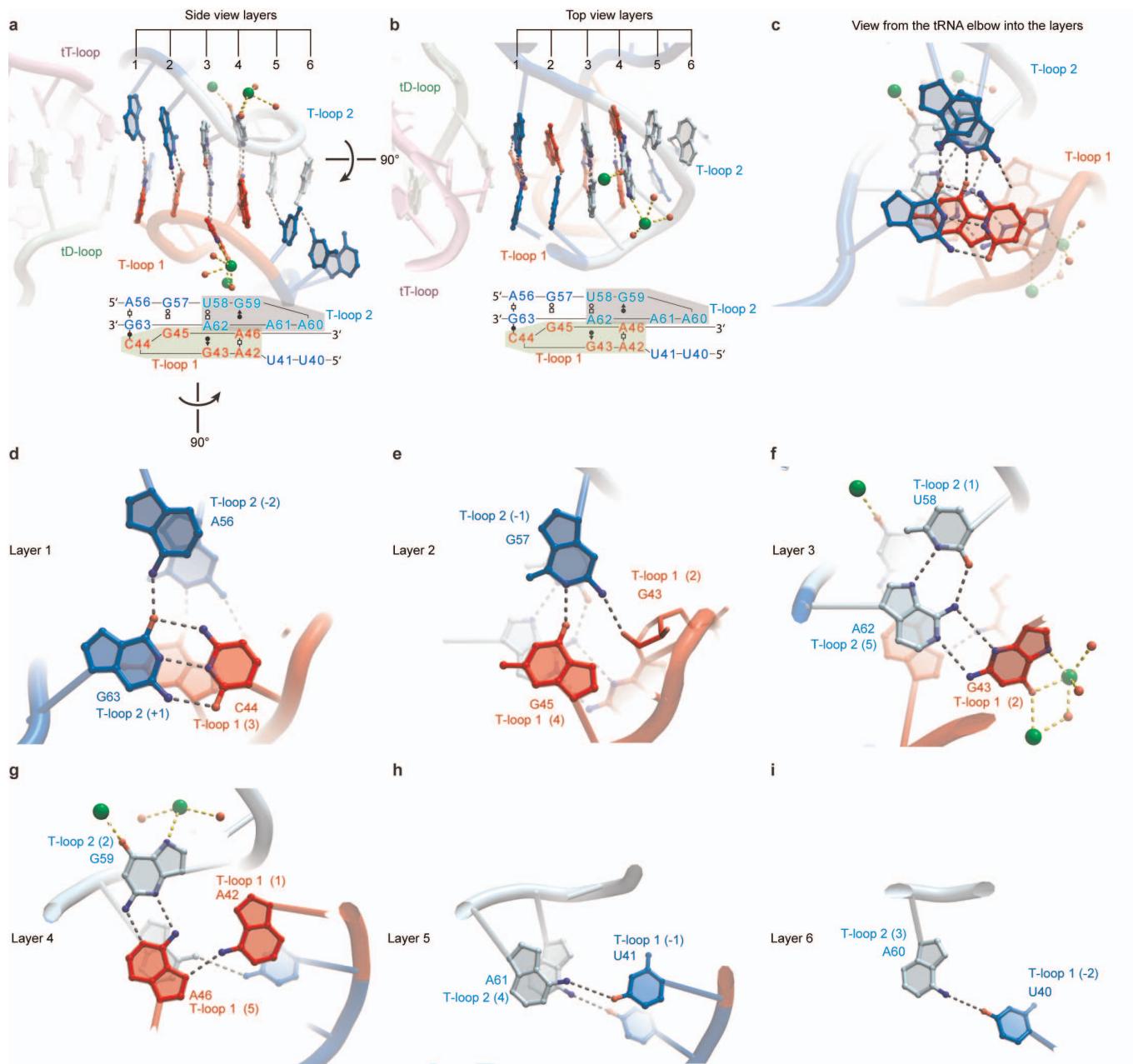


Extended Data Figure 3 | Intermolecular interface of the stem-I-tRNA complex and phylogenetic conservation. **a**, The residue 29–30 bulge allows the *O. iheyensis* stem I to bend and track the tRNA backbone. **b**, Structure²¹ of the equivalent region of the *Geobacillus kaustophilus* T-box, which folds as a C-loop⁵². **c**, Solvent-accessible surface coloured according to area buried from

blue or white (no burial) to red ($>75 \text{ \AA}^2$ per residue). **d**, Open-book view. **e**, Surfaces coloured by phylogenetic conservation of stem I residues (see Extended Data Fig. 1). Conservation of tRNA is not indicated. **f**, Open-book view. For comparison, RNase P binding to pre-tRNA buries⁶ $2,959 \text{ \AA}^2$.

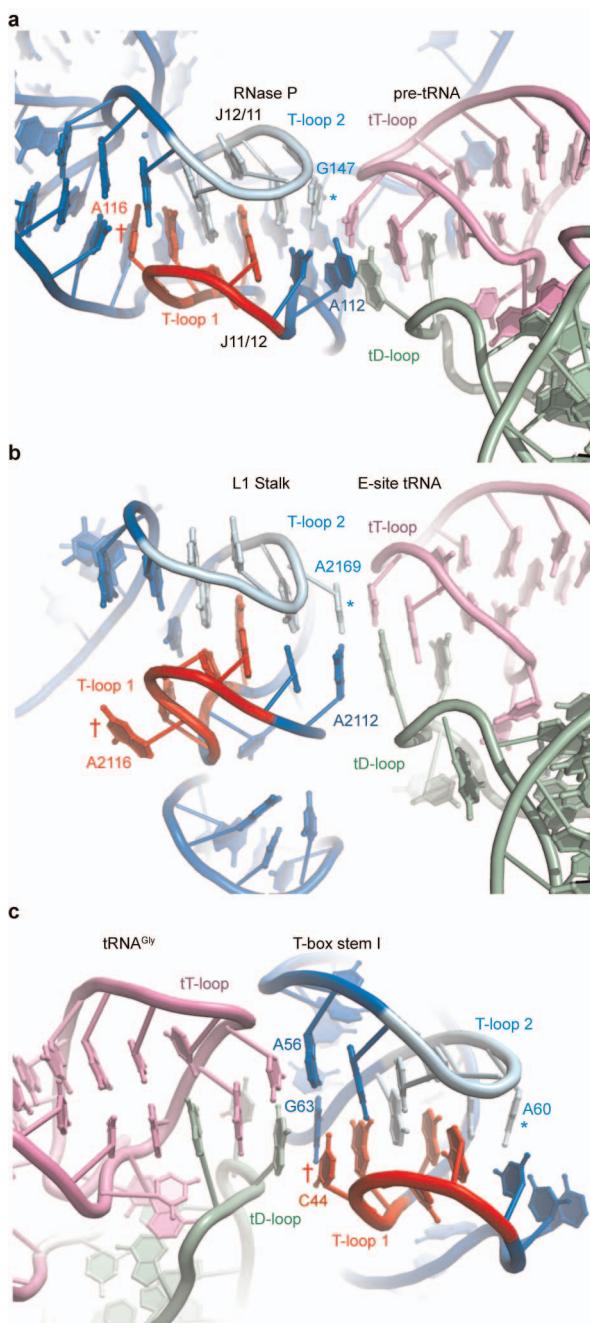


Extended Data Figure 4 | T-box can accommodate bulky post-transcriptional modifications at tRNA position 37. A86 of the T-box stem I is unstacked from the duplex formed by pairing of the tRNA anticodon and the T-box specifier by its participation in the sheared A•A pair at the base of the loop E motif. This results in a pocket (dotted line) that could accommodate the large, modified nucleobases (for example, the three-ring heterocycle wybutine in yeast tRNA^{Phe}) often present at position 37 of tRNAs (Fig. 2a). Note the single hydrogen bond between tU32 and tA38.

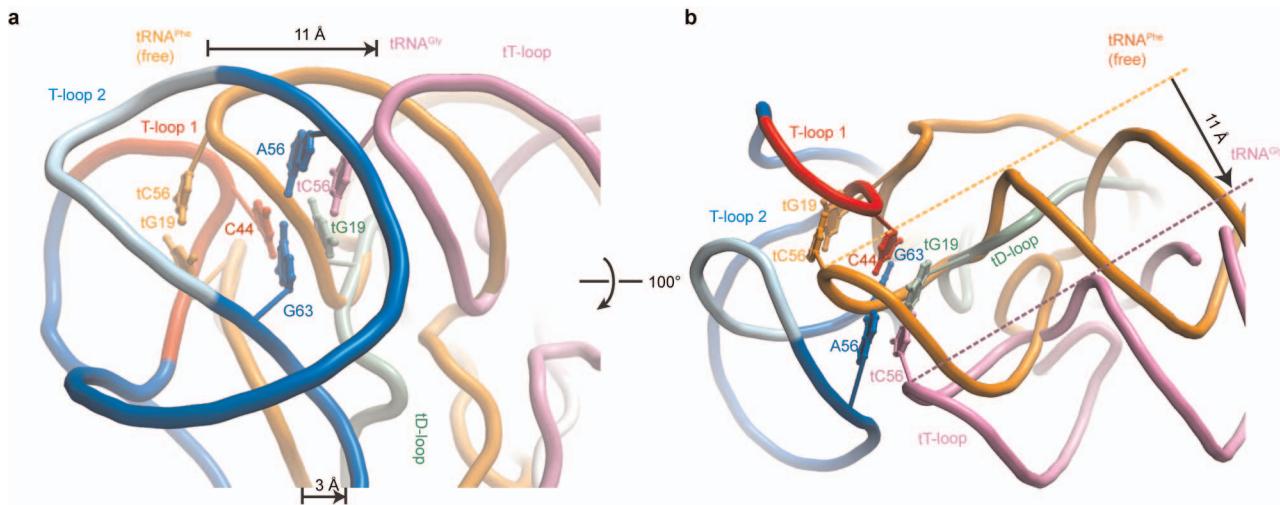


Extended Data Figure 5 | Structural organization of the interdigitated T-loop motifs of the T-box stem I. **a**, Six base-paired layers form the stem I distal domain core (view as in Fig. 2b). Dashed black lines denote base–base hydrogen bonding. Dashed yellow lines denote Sr^{2+} -ion–water (green and red spheres, respectively) interactions. Leontis–Westhof symbols³⁰ depict

non-canonical pairing. **b**, Orthogonal view. **c**, View from the tRNA elbow. **d–i**, Detail of layers 1–6. The position of each nucleotide within the pentaloop (1–5) or outside of it (negative and positive denote 5' and 3' directions, respectively) is in parentheses (Supplementary Discussion).

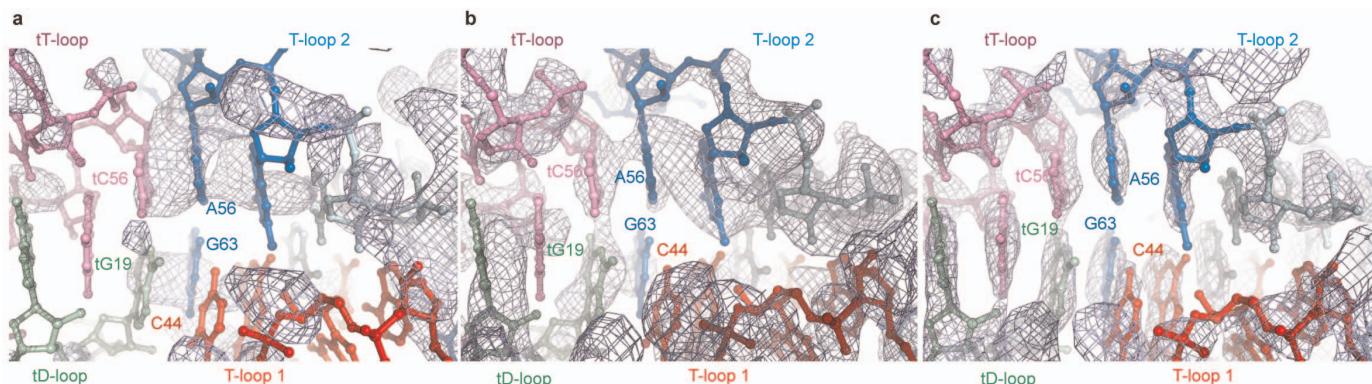


Extended Data Figure 6 | Comparison of the interdigitated T-loops of RNase P, the L1 stalk and the T-box. **a, b,** RNase P (a) and the L1 stalk (b) use equivalent surfaces formed by two purines—namely, nucleotide -2 of T-loop 1 and nucleotide 3 of T-loop 2 (asterisk)—to stack on the elbows of pre-tRNA and the E-site tRNA, respectively. **c,** By contrast, the T-box uses the opposite surface, using a triple (Fig. 2c) formed by nucleotide 3 of T-loop 1 (dagger symbol), nucleotide -2 from T-loop 2, and nucleotide $+1$ from T-loop 2 to stack on the tRNA elbow.



Extended Data Figure 7 | Displacement of the tRNA elbow induced by T-box binding. Although the anticodon stem of the free tRNA^{Phe} (orange) superimposes well on that of the T-box-bound tRNA^{Gly} (T-loop and acceptor stem in violet; D-loop and ASL in pale green; Fig. 3b, c), the elbow (tG19–tC56)

clashes with the stem I distal domain. T-box binding induces bending about the t26•t44 hinge (Fig. 3c) that displaces the tRNA elbow and acceptor stem away from the distal region of stem I by 11 Å (arrows), allowing the tRNA elbow to associate with the T-box. **a**, Side view. **b**, Top view.



Extended Data Figure 8 | Representative electron density maps. Portions of electron density maps corresponding to the interdigitated stem I T-loops and the tRNA elbow (oriented like Fig. 2b). **a**, Unbiased, 3.2-Å-resolution density-modified SAD electron density (1.0 s.d. above mean peak height) shown as grey mesh superimposed on the refined model. **b**, Electron density resulting from

combination of SAD and molecular replacement (using a search model consisting of two copies each of tRNA^{Phe} and a K-turn in complex with YbxF; Methods) phases followed by density modification (1.5 s.d.) **c**, Composite simulated anneal-omit $2|F_o| - |F_c|$ electron density calculated using the final model (2.0 s.d.).

Extended Data Table 1 | Thermodynamic parameters from ITC analysis

T-box *	tRNA ^{Gly}	K _d (μM)	log[K _a]	ΔH (kcal mol ⁻¹)	-TΔS (kcal mol ⁻¹) †
14-182 (wt) ‡	wt	0.209 ± 0.025	6.68	-29.9	21.0
14-158 (wt) §	wt	0.194 ± 0.055	6.72	-41.2	32.2
14-113 (wt)	wt	0.150 ± 0.018	6.82	-25.1	15.9
14-113del_kturn ¶	wt	0.097 ± 0.008	7.01	-23.6	14.2
Δ29-30 (Δ42-43)	wt	1.84 ± 0.42	5.74	-27.0	19.3
G43U (G55U)	wt	1.68 ± 0.16	5.78	-39.7	31.9
C44U (C56U)	wt	8.20 ± 2.9	5.08	-45.3	38.5
C44A (C56A)	wt	0.307 ± 0.038	6.51	-30.7	21.9
C44G (C56G)	wt	3.4 ± 2.3	5.47	-42.1	34.7
A42G/A46G (A54G/A58G)	wt	9.0 ± 2.1	5.05	-30.0	23.2
G56U (G68U)	wt	4.6 ± 0.44	5.34	-32.0	24.8
G56A (G68A)	wt	0.406 ± 0.060	6.39	-33.2	24.6
G59U (G71U)	wt	0.538 ± 0.12	6.27	-45.4	36.9
G59A (G71A)	wt	1.69 ± 0.30	5.77	-43.0	35.2
A61U (A73U)	wt	0.125 ± 0.019	6.90	-35.4	26.1
U58G/A62G (U70G/A74G)	wt	12.6 ± 8.3	4.90	-28.2	21.6
G63C (G75C)	wt	8.7 ± 0.84	5.06	-30.0	23.2
C44G/G63C (C56G/G75C)	wt	1.9 ± 1.5	5.72	-39.8	32.2
C44A/G63A (C56A/G75A)	wt	0.433 ± 0.31	6.36	-32.0	23.5
G87U (G99U)	tC36A	7.6 ± 4.2	5.12	-17.2	10.4
A90C (A102C)	wt	0.641 ± 0.167	6.19	-26.9	18.6
A90G (A102G)	wt	0.070 ± 0.012	7.15	-31.5	21.9
14-113 (wt)	tA37U	0.125 ± 0.018	6.90	-35.4	26.1
14-113 (wt)	3'GAAA #	0.083 ± 0.016	7.08	-31.8	22.3

* Residue numbers are for *O. iheyensis* glyQ T-box. Structurally equivalent *B. subtilis* glyQS T-box numbers are in parentheses.

† All experiments were performed at 293 K.

‡ This experiment corresponds to the titration shown in Fig. 1a.

§ This experiment corresponds to the titration shown in Fig. 1b.

|| This experiment corresponds to the titration shown in Fig. 1c and Extended Data Fig. 2a.

¶ This experiment corresponds to the titration shown in Extended Data Fig. 2b.

Engineered crystallization construct.

Extended Data Table 2 | Data collection and refinement statistics

	Crystal I	Crystal II	Crystal III	Crystals I-II-III
Data collection				
Space group	C222 ₁	C222 ₁	C222 ₁	C222 ₁
Cell dimensions				
<i>a, b, c</i> (Å)	100.3, 108.4, 266.8	99.1, 107.5, 266.1	100.3, 108.3, 267.8	100.0, 108.1, 266.9
α, β, γ (°)	90, 90, 90	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	28.42-3.20 (3.31-3.20)*	133.1-3.87 (4.08-3.87)	133.9-3.76 (3.97-3.76)	134-3.43 (3.61-3.43)
R_{merge} (%)	6.2 (104.8)	6.2 (67.1)	6.7 (59.5)	10.8 (82.8)
$<\!>/<\!\sigma(\!>)$	11.1 (1.5)	14.3 (2.2)	14.2 (2.3)	23.7 (2.3)
Completeness (%)	98.9 (98.8)	99.8 (99.7)	99.4 (99.2)	99.3 (96.0)
Redundancy	6.1 (6.2)	4.8 (4.8)	4.8 (4.7)	12.8 (6.2)
Refinement				
Resolution (Å)	28.42-3.20 (3.31-3.20)			
No. reflections	24325 (2369)			
$R_{\text{work}}/R_{\text{free}}$ (%)	19.6 (33.6) / 25.2 (38.2)			
No. atoms	8742			
RNA	7586			
Protein	1048			
Ion	86			
Water	22			
Mean <i>B</i> -factors (Å ²)	128.0			
RNA	118.7			
Protein	192.7			
Ligand/ion	176.4			
Water	64.2			
R.m.s. deviations				
Bond lengths (Å)	0.001			
Bond angles (°)	0.37			

One crystal was used for each of the three data sets.

* Highest resolution shell is shown in parenthesis.

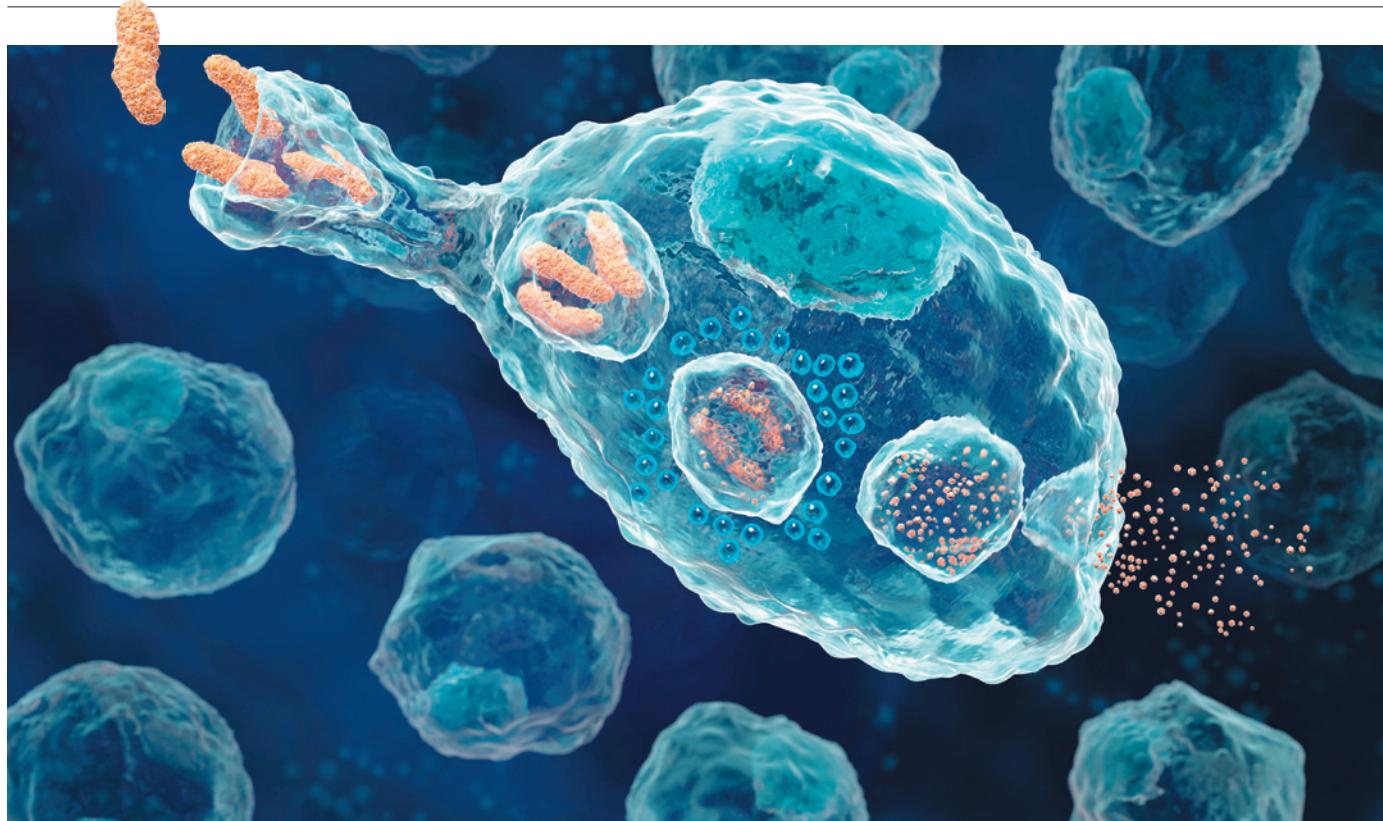
CAREERS

COLUMN It is not easy to let go of a scientific identity after leaving the lab p.369

HIRING US online science recruiting continues to grow in 2013 p.369

 **NATUREJOBS** For the latest career listings and advice www.naturejobs.com

DAVID MACK/GETTY



A macrophage white blood cell (centre) engulfs and destroys bacteria (orange) and spews out the remnants.

LIFE SCIENCES

Industrial immunology

In a tough job market, immunologists are in demand. To move from academia to a biotechnology or drug company, researchers must explore the options and stay flexible.

BY ALLA KATSNELSON

Olivia Schneider realized early in her graduate work on immune-cell signalling that she had no interest in becoming an academic researcher. “I didn’t want to work in a lab, or to write grants,” she says. In 2009, when she finished her PhD at the University of Cincinnati in Ohio, the global recession was in full swing and employment options looked scarce. Her husband had a well-paid job in the area, so relocating was impractical. Then Schneider saw an advert for part-time work through a local contract-research organization, doing tissue culture and cloning for a recombinant-protein manufacturer called Shenandoah

Biotechnology in Warwick, Pennsylvania. “I just wanted to get my foot into biotech in some regard,” she says. “I took this position — that I was way overqualified for — with the hope that it would turn into something else.”

And it did. The project involved using antibody-based kits to test the quality of Shenandoah’s products. Flying back and forth to give talks on progress, Schneider developed a rapport with Shenandoah’s owners. When the project ended unsuccessfully, they invited her to join their firm, working remotely as an associate director of research and development. Fifteen months later, she became chief scientific officer.

Schneider now brainstorms new products on the basis of feedback from researchers. She

manages quality control, represents the company at trade shows and conferences, trains sales personnel and even tackles marketing and customer relations. The variety suits her. “I try to tell people, try to pick out what you liked in graduate school. There is a job associated with that,” she says. As a student, Schneider had loved giving presentations and talking to people about their own work and science in general, but had little interest in actually analysing data. “I kind of created a job that reflects that,” she says.

Industrial opportunities for early-career scientists have decreased in the wake of the recession, but in some quarters, immunology has done better — especially when compared to fields such as neuroscience. Immunology ▶

► and oncology are, for example, the most active fields for recruitment at the executive level, says Kenneth Israel, founder of Search Group of San Diego, a recruitment firm in California that focuses on executive-level positions in the life sciences.

The interest may be fuelled by companies increasing investment in biological drugs such as antibodies, says Ravi Chari, executive director of chemistry and biochemistry at ImmunoGen, in Waltham, Massachusetts, which develops antibodies for treating cancer. "Even traditional pharma companies like Merck and Eli Lilly are now getting into immunology," he says. Jonathan Deane, a researcher at the Genomics Institute of the Novartis Research Foundation in San Diego, notes that several immunomodulating drugs have been approved in the United States in recent years, and there has been a stream of positive news about immune-based cancer treatments. Still, the job market is affected by broader industry forces, he warns. "Companies have their budgets — that's just global," he says, noting that many are making cuts.

CLEAR GOALS

In the current climate, immunologists need to be clear about the kind of job they are looking for. One niche is research positions, either at drug-development companies or at firms developing reagents and research materials for use by academic or industry scientists. But there are options beyond the bench, too — such as medical-liaison officers, who cultivate contacts in a particular therapeutic field to create awareness of a product or company.

For research positions, PhD holders can expect annual starting salaries ranging from €45,000 (US\$59,000) to €65,000 in Europe, says David Marimuthu, senior recruitment consultant at SIRE Life Sciences, a staffing company based in Amsterdam. However, he adds, someone with a few years of industry experience or specialized knowledge in a niche area could get double that — or even more

with management experience. Starting research salaries are closer to about US\$80,000 in the United States, says Phil Mericantante, co-owner of Adante Staffing in Woburn, Massachusetts, which specializes in the life sciences. Medical-liaison officers might start at between \$85,000 and \$100,000 depending on the size of the company, and regulatory positions bring in something like \$50,000–60,000, says Mericantante.

Recruiters do not tend to work with



Ravi Chari says the best applicants have explored several areas of research.

entry-level applicants, because companies are not generally interested in paying finders' fees for such jobs, says Marimuthu. But it is a different story for people with highly specialized research backgrounds. "Companies developing one product, if they find a real genius in their field, they are willing to pay," he says, and recruiters may be useful.

As with most career transitions, networking is a tried and true way to learn about the types of opportunities available and to land a job. Job seekers should do more than just show up to events, advises Schneider. They should seek out people from interesting companies at conferences, and mine their university career centre's alumni list for contacts. Applicants can request informational interviews to find out about the company, and then ask who else they should speak to. "Once you know you want to go after a particular career, you really have to do your homework," says Michael Crowley, director of business development at the biotechnology firm Genentech in South San Francisco, California. "You have to know the company, and you have to bring really good energy to the interviews. If interviewers are not sensing the investment, they will just write you off."

As an undergraduate, Crowley was "devoutly focused on immunology" but, like Schneider, he realized during his doctorate that he did not picture himself running an academic lab. In 1999, a year after finishing his PhD on immune cells called $\gamma\delta$ T cells at Stanford University in California, he fell into a consulting job at the now-defunct Cellegy Pharmaceuticals. He loved the operations side of drug development, so he studied for a business degree, and Genentech hired him in 2003, after a summer internship. Now he leads a three-person team that seeks out research collaborations and licensing deals relating to immunology and infectious diseases.

Business-development positions differ depending on the size of the company, Crowley notes. At Genentech, he focuses on what he calls the "buy-side" — bringing in intellectual property on the basis of a wish list provided by company researchers. For example, in February, his group signed a deal with a San Diego startup that had found an innovative way to pursue a target of interest. "Small companies are the sellers, and that's a different kind of job," he says.

Crowley adds that he does not feel as if he abandoned the work he loves to go into business. "I'm still very close to the science here," he says. "But it's also about communication and relationships, which I felt I couldn't get at the bench." Job seekers should decide from the start whether they want to remain in the lab or forge a non-research career, he says. "If you're ambivalent, it will hold you back in both realms."

Those who do want to stay at the bench should attempt to broaden out beyond a single field of expertise, says Chari. He has a PhD in organic chemistry but did his postdoctoral work in protein chemistry, enzymology and related fields. His broad background helped him to get hired at ImmunoGen in the late 1980s.

DIFFERENT STROKES

Company cultures take some getting used to. Probably the biggest difference from academia is team involvement, says Deane. Developing a drug involves multiple teams working in concert, he says: one group might be doing the immunology, but at the same time a second group is looking at protein expression, a third at X-ray crystallography and a fourth at pharmacodynamics. Each group must stay in close communication with the others, and with managers guiding the project.

Then there is the issue of publishing research. Some companies encourage it, whereas others prohibit it. Kerry Casey, a research scientist at MedImmune in Gaithersburg, Maryland, the biological-drugs research and development arm of AstraZeneca, feels fortunate to be at a company that promotes publishing. In her two years there, she has worked on a high-priority project in respiratory inflammation and immunity, and she currently has two manuscripts in the works. MedImmune is close to the US National Institutes of Health (NIH) in Bethesda, Maryland, so Casey can easily pop over for conferences. Many of her co-workers once did research at the NIH, and still have colleagues there. "We are really linked into the academic community," she says.

Most industrial researchers work under some corporate limitations, however. "As a university professor, you are your own small business and your own chief executive," says Lewis Lanier, a microbiologist and immunologist at the University of California, San Francisco. "Companies are dictatorships — unless you're the chief executive, you're not in charge." ■

Alla Katsnelson is a freelance writer in Northampton, Massachusetts.



"We are really linked into the academic community."

Kerry Casey

COLUMN

Still a scientist

It is not easy to let go of a scientific identity after leaving the lab — nor is it necessary, says **Chandrika Nair**.

If I choose a career outside the lab, will I still be a scientist? That question often occupies my thoughts as I consider my options post-PhD.

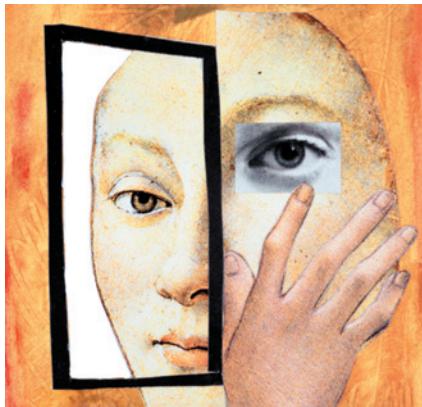
Imagining a future outside research is not easy: I have become accustomed to this environment. But it makes sense, given that more than half of current science PhD students will leave the lab and spread their wings elsewhere (see *Nature* 475, 533–535; 2011). Perhaps a science-related job could be a workable compromise.

Scientific research relies on a large network of support, from teachers to research councils and charities. There are also editors, journalists and communicators, policy-makers and campaigners. A job in any of these areas could tick the science box — keeping up with research trends as a journalist, for example, or reviewing evidence as a policy-maker.

But would I want to be the person facilitating science rather than doing it myself? Should I exchange my lab coat for a suit? My primary motivation for considering other career paths is that experimental work can be lonely. As I sit pipetting at a fume hood or staring at cells in a darkened room, I admit, I crave human interaction. Still, academic research has many advantages, and the one I treasure most is freedom. Many researchers have flexible schedules and intellectual independence. And the idea of working on basic ‘blue skies’ research appeals to me. Few other jobs pay their employees to indulge their child-like curiosity, to pursue projects that may or may not succeed.

If I do leave the bench, will I forfeit my identity as a scientist? Can a person be a scientist and not work in a lab? The scientist in me decided to address this deep, dark question by gathering some data. I surveyed nine people working in science-related fields, including a medical charity, a learned society and a research-funding body. The results? Although they no longer consider themselves scientists, all respondents said that their science background has stayed with them. That knowledge and training are relevant to their job on a scale varying from ‘occasionally’ to ‘routinely’.

But what interested me most was the consensus that it is important to have a scientific outlook. The people I asked said that their training affects how they approach problems at work, encouraging them to seek evidence and test ideas. They have exported aspects



of the scientific method to real-life settings. One person said that they typically question and research sources of information more than their colleagues who do not have scientific backgrounds.

I do not know what career I will be pursuing in several years’ time. One highly desirable scenario would be presenting exciting findings at a conference with several high-impact papers under my belt, maybe a grant or two, and a permanent position within reach. However, I am coming to the pleasant realization that I could easily find fulfilment in other occupations, especially if I feel that I am putting my scientific training to good use.

My very preliminary data suggest that people who have left research might not fit the conventional definition of a scientist, but still retain the identity of scientists by training. As someone who often wishes that important aspects of our lives — from advertising to health care and politics — were more evidence-based, I welcome the possibility of science-literate people having diverse roles in society. Perhaps I will even still be able to call myself “a scientist”. ■

Chandrika Nair is a PhD student in microbiology at Imperial College London.

CORRECTION

The Careers Feature ‘Middle Eastern promise’ (*Nature* 500, 111–112) wrongly stated that the New York University Abu Dhabi offers master’s-level courses. It does not.

HIRING

US job ads increase

There were 6% more advertisements for US science-related positions posted online in June 2013 than in June 2012, following a 22% increase from 2011 to 2012, according to an analysis of 25,000 recruitment sources including LinkedIn and CareerBuilder. The study, released on 31 July by Wanted Analytics in Quebec City, Canada, also found that the number of employers advertising jobs was up slightly in 2013 after a 15.8% increase from 2011 to 2012. The number of academic posts advertised, however, was down 16.5% from 2012. The biggest increases in recruitment were in San Francisco and Los Angeles, California. In both areas, demand rose by almost one-fifth, owing in part to increased hiring at the University of Southern California and at biomedical research firms including Genentech.

DATA-SHARING

Private beats public

Researchers in some fields are more likely to share data in response to personal requests than by routinely submitting to public repositories, says a study (J. C. Wallis *et al.* *PLoS ONE* 8, e67332; 2013). The authors conducted 43 interviews with researchers who use wireless sensing, including biologists and seismologists; in more than two-thirds, participants did not report submitting to repositories, sometimes because a suitable one did not exist or because they wanted to know how data would be used. Co-author Jillian Wallis, an information-studies researcher at the University of California, Los Angeles, emphasizes that using repositories makes data easier to find and increases citations.

ACADEMIA

US health premiums rise

More than one-quarter of US academic institutions said in a survey that they had raised the costs of health-care premiums for faculty members and staff this year in preparation for the Affordable Care Act, which is set to ramp up in 2014. The 2013 *Employee Health Benefits Survey of Higher Education* by the College and University Professional Association for Human Resources in Knoxville, Tennessee, found that premiums rose by an average of 3.25% for employee-only plans and 5% for employee-family plans. Two-thirds of respondents said that they are still trying to determine which adjunct or part-time faculty members are eligible for benefits.

TIME HEALS ALL WOUNDS

Seconds out.

BY GRACE TANG

Hannah's hair glowed amber in the porch light as I watched her through the semicircular pane of frosted glass. In the next second, the front door would click shut, cutting her off from my life. I willed myself to rise, to go after her and stop her, but movement was something that belonged to the next second, and the next second would not come.

Time had, as expected in my condition, slowed dramatically, and in this state I observed the world in excruciating detail: the scraps of 'meat' that had fallen from our TV-dinner trays were spreading pools of oil on the blue paper of the unpaid hospital bills. The unkempt pile lay strewn across the coffee table, and in the centre of the heap lay a half-carat diamond ring, the band scratched and faded from years of wear.

Two decades ago, on the last day that time still flowed at a regular pace regardless of my emotional state, the ring had been new and unblemished, sitting inside a ring box on my doctor's desk.

"I'm going to return it today," I choked, my forehead resting on my palm, trying desperately to hide my emotions, "it's not fair to Hannah."

Doctor Kwon sat opposite me in silence. I knew he wanted to tell me that he could up the dose of the medication again, and that the voices and hallucinations would stop, allowing me to lead a normal life. I knew he wanted to say that my disease would not gradually chip away at my personality till all that was left of me was a raving lunatic who only externally resembled my former self.

But all Doctor Kwon could do was put his hand on my shaking one. He hesitated.

"Dan, there is something that just may work."

I looked up, hopeful.

"Now don't get too excited. It's an experimental procedure, a brain surgery, never tried before in humans. A team in this hospital was going to start looking for volunteers."

I snatched the ring off the table and pushed it back into my jacket pocket.

"Sign me up."

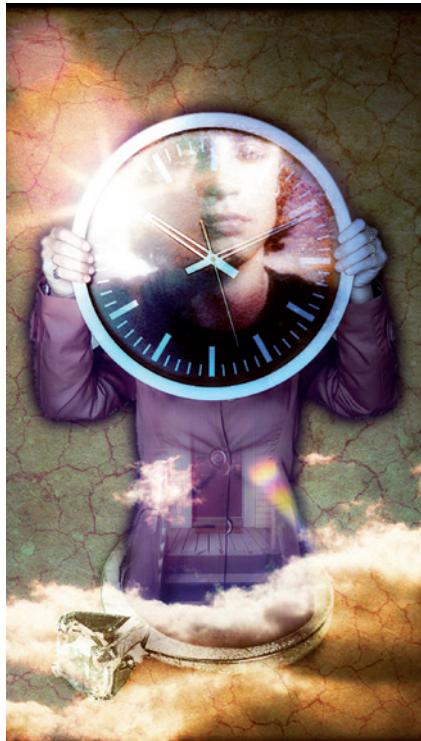
At first it seemed as though my hallucinations had stopped after the surgery, but perhaps it was just that the nature of my hallucinations had changed. My first hot bath since the

NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm



operation was glorious. I soaked in the water for what felt like hours, until I was worried Hannah would start to wonder if I had passed out. Hannah eyed me with concern when I emerged, dripping.

"Something wrong, dear? Did it hurt?" She gingerly caressed my head near the site of the surgery.

"No, that felt great."

"Then why did you come out right away?"

All too soon I was back in Doctor Kwon's office. The team that had operated on me stood by, fascinated. He held a stopwatch in one hand and offered me a piece of gum from a bowl in the other. Under his instructions, I put the gum in my mouth. I enjoyed the mint flavour immensely, but he remained silent, watching me eat until it became awkward to continue.

"Uh, should I keep going?"

He clicked the stopwatch off.

"How long have you been eating?"

"I don't know, five minutes?"

"Fascinating..."

"What?"

Out of nowhere, he hit the side of my arm. It stung.

"Ow! What was that for?"

Again, he stood there dumbly for the longest time before responding.

"Sorry. How long did the pain last when I hit you?"

"Three minutes?"

"It seems that a side effect of the surgery is that emotional responses slow your perception of time. I gave you that piece of gum no more than two minutes ago..."

It was trivial in the beginning — seconds could stretch into minutes — but my condition gradually got worse. After a year, a few seconds could drag like hours when I got emotional. It got to the point where I could no longer tell time enough to function. I lost my job, and my main source of income was the chump change time-perception researchers threw at me when they needed a guinea pig.

Tonight, Hannah had come back from her second job exhausted.

"Dan," she sighed, putting her face in her hands.

I could sense the distress in her voice, which alone was enough to trigger enough of an emotional response that time began to slow.

"Yes... deaaar?"

Like a drunk at a party, vaguely aware of my slurred speech, I struggled to speak at a normal rate, unsure if I was succeeding.

She inhaled sharply, her voice cracking.

"I'm sorry..."

She took off her ring for the first time in 20 years. Time crawled slower than it ever had, nearly still. Was my condition deteriorating, or had I never felt such a strong emotion in my life? The diamond sparkled like a tiny glacier as it drifted through the ocean of space between her hand and the top of the coffee table.

Now the halo-like rays from Hannah's blonde crown faded like the setting sun as she descended the front stairs of the porch.

The wall clock ticked, deafening. With the arrival of the new second, the door clicked into place, but I had not moved an inch. Hannah disappeared from view. My eyelids fell, the first stage of the 300-millisecond sequence of an eye blink. The flow of air from my lungs slowed as I reached the nadir of my breath. If the Universe would progress another hundred milliseconds, my eyes would open again, and I would take my next breath. I waited and waited, but neither happened. I knew, like they said, that time would heal this wound, but time had ceased to flow. ■

Grace Tang is a graduate student in psychology at Stanford University.